numero 164
aprile 1982
anno XV
volume XXVIII

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

# Effetti biologici di radiazioni ionizzanti di basso livello

*I dati fino a oggi raccolti indicano che, in confronto con altri, il rischio per l'uomo associato alle onnipresenti radiazioni di basso livello da sorgenti naturali e artificiali è molto piccolo*

di Arthur C. Upton

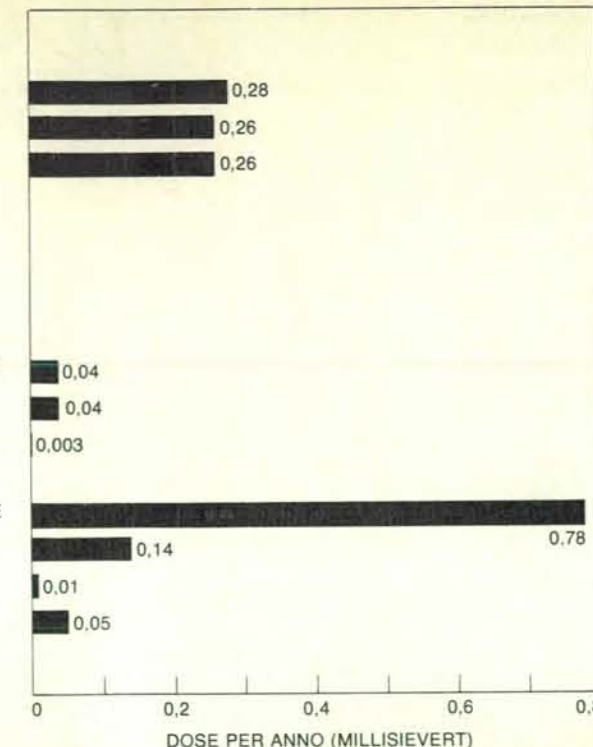
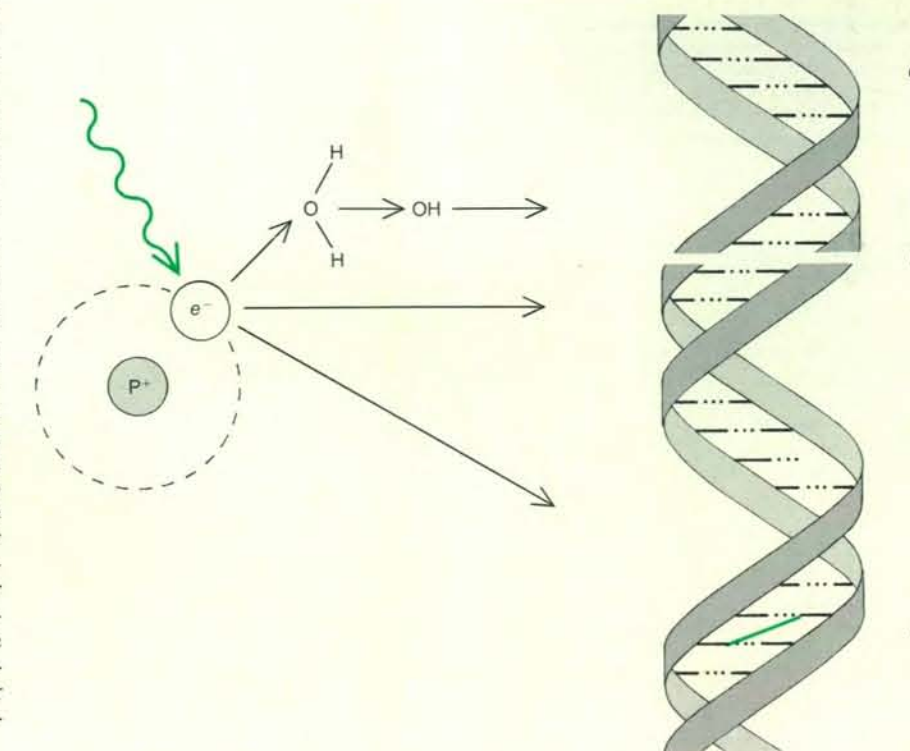
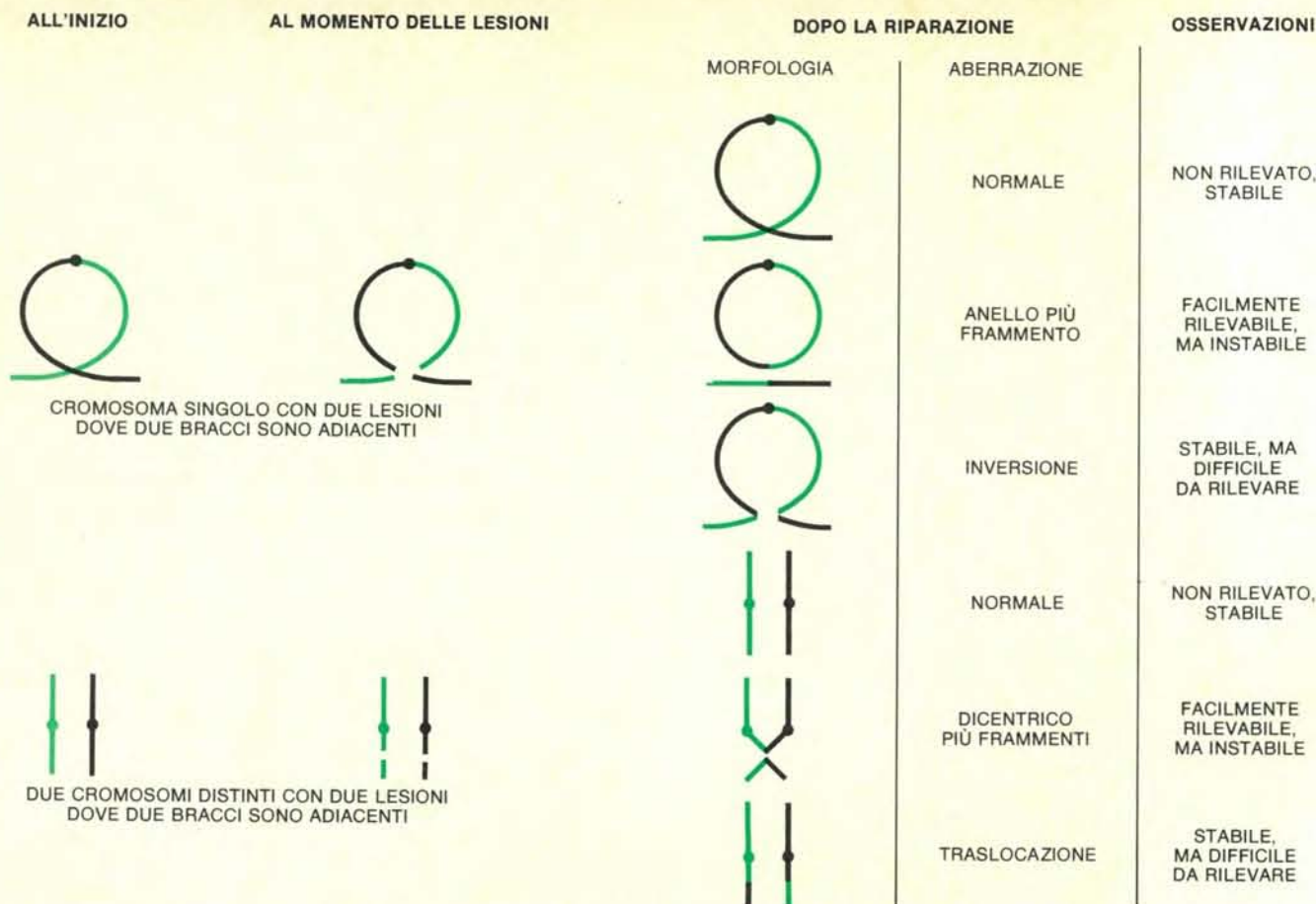
Per tutta la vita ciascuno di noi è immerso in un mare di radiazioni ionizzanti di basso livello, emesse da sorgenti naturali. Per una persona media, l'esposizione lungo tutto l'arco della vita raddoppia a causa delle fonti di origine umana, in particolare quelle mediche. La radiazione può causare neoplasie e difetti genetici; si pone quindi il problema di stabilire quanto sia dannosa la radiazione di basso livello sia essa di fondo o artificiale (prodotta da sorgenti costruite dall'uomo). Il problema è intensamente allo studio da anni e le prove raccolte fin qui indicano che il rischio, se esiste, è troppo piccolo per essere rilevato; pertanto, rispetto ad altri rischi a cui siamo regolarmente esposti, il rischio complessivo dovuto alla radiazione ionizzante di basso livello non sembrerebbe destare alcun motivo di preoccupazione.

Vi è un altro problema, collegato al primo, e relativo alle soglie. L'effetto della radiazione ionizzante di basso livello è cumulativo, o la radiazione è dannosa solo quando supera un ben preciso livello di soglia? A questo proposito i dati di fatto sono più ambigui. Per taluni tipi di radiazioni e per certi effetti, è chiaro che l'influenza della radiazione è proprio cumulativa. In altri casi, però, sembra più plausibile ipotizzare l'esistenza di un valore di soglia.

A differenza di altre forme di energia radiante, la radiazione ionizzante può fornire una quantità di energia ben localizzata, a un materiale che lo assorbe, sufficiente per ionizzare atomi e molecole lungo il suo percorso (in modo tale, cioè, da togliere o aggiungere loro elettroni). Alcune radiazioni ionizzanti sono di natura elettromagnetica e alcune sono formate da particelle subatomiche caratterizzate da vari valori di massa e di carica. Fra le radiazioni ionizzanti elettromagnetiche vanno inclusi i raggi X e i raggi gamma, che sono caratterizzati da lunghezze d'onda relativamente corte e da elevata frequenza. Fra le radiazioni corpuscolate si trovano invece elettroni, protoni, neutroni e particelle alfa.

Esistono diverse unità di misura per le quantità e le dosi di radiazioni ionizzanti. L'unità più vecchia è il roentgen, che è una misura della quantità di ionizzazione indotta nell'aria. Le unità principali usate per esprimere le dosi assorbite da materia vivente sono il rad (un rad è pari a 100 erg per grammo di tessuto) e il gray (un gray è pari a un joule per chilogrammo di tessuto, cioè a 100 rad). Dato che a parità di dose di solito le radiazioni corpuscolate causano più danni dei raggi X o gamma, sono state introdotte le unità rem e sievert, in modo da poter confrontare le dosi di tipi diversi di radiazione, rispetto ai loro effetti biologici. In termini un po' approssimativi, un rem è la quantità di radiazione che produce un effetto biologico equivalente a quello prodotto da un rad di raggi gamma. Analogamente, un sievert è la quantità di radiazione il cui effetto biologico è pari a quello prodotto da un gray di raggi gamma; un siervert è pari, quindi, a 100 rem. Le unità usate per esprimere dosi collettive sono il rem-persona e il sievert-persona: si ottengono per moltiplicazione, moltiplicando cioè la dose media per persona per il numero delle persone esposte. Nel caso di una popolazione esposta di 100 persone, un rem per ciascuna persona è pari a 100 rem-persona, ovvero a un sievert-persona.

La radiazione naturale di fondo è composta principalmente da (1) raggi cosmici, (2) emissioni dovute alla disintegrazione di uranio, torio, radio e altri elementi radioattivi nella crosta della Terra e (3) emissioni da potassio 40, carbonio 14 e altri isotopi radioattivi presenti in natura negli organismi. Collettivamente la dose media ricevuta dagli organi

**L'energia della radiazione in arrivo su una cellula vivente viene dissipata in modi diversi a seconda della natura della radiazione. In genere i raggi X e i raggi gamma hanno un basso tasso di trasferimento lineare di energia: penetrano in profondità, ma producono ioni solo qua e là lungo il loro percorso. Le radiazioni corpuscolate, come elettroni e particelle alfa, invece, presentano un tasso elevato di trasferimento lineare di energia. Gli schemi qui sopra raffigurano i due tipi di radiazione nel caso di dosi uguali di raggi gamma *(a)* e di una radiazione corpuscolata *(b)*. I puntini rappresentano eventi di ionizzazione provocati dalla dissipazione di energia lungo i percorsi della radiazione *(linee in colore)*. In ambedue i casi si verificano piccole zone con elevata concentrazione di eventi di dissipazione di energia *(cerchietti in colore)*, ma questi casi sono meno frequenti, per dose unitaria, in *a* che in *b*. L'illustrazione è basata sulle ricerche di G. W. Barendsen nei Paesi Bassi.**

**È qui rappresentata in millisievert l'esposizione alla radiazione di un tipico cittadino degli USA, da sorgenti naturali e artificiali. Un sievert è la quantità di radiazione con effetti biologici pari a quelli di 100 rad (un rad è pari a cento erg per grammo di tessuto) di raggi gamma. L'intensità media totale di dose per anno è di 0,8 millisievert da sorgenti naturali e 1,06 millisievert da sorgenti artificiali. Una radiografia odontoiatrica trasferisce alla guancia circa un millisiervert.**

**L'effetto di maggiore gravità della radiazione di basso livello è un danno al DNA. L'effetto può essere diretto o indiretto. A sinistra nell'illustrazione si vede un atomo di idrogeno, costituito da un protone e da un elettrone, che viene ionizzato. (L'elettrone viene spostato quando l'atomo assorbe un fotone di radiazione.) Nel caso di un effetto indiretto *(freccia in alto)* un elettrone secondario interagisce con una molecola di acqua ($H_2O$) e dà così luogo a un radicale libero (OH), che provoca il danno al DNA. Nel caso dell'effetto diretto *(frecce inferiori)* l'elettrone stesso interagisce con il DNA. La rappresentazione schematica del DNA a destra mostra un segmento normale della molecola *(a)* e tre fra i vari tipi di danno che possono essere causati da effetti diretti o indiretti di ionizzazione: la rottura di un doppio filamento nella doppia elica del DNA *(b)*, la cancellazione di una base *(c)* e il legame chimico trasversale fra due filamenti di DNA *(d)*.**

**Spesso i danni causati ai cromosomi dalla radiazione vengono direttamente riparati dall'organismo, ma talvolta la riparazione avviene in modo errato e l'esito è una aberrazione cromosomica. In questa illustrazione sono raffigurati i casi di un singolo cromosoma *(in alto)* con una lesione indotta da radiazione in ciascun braccio e di due distinti cromosomi *(in basso)*, ciascuno dei quali presenta una lesione in un braccio. Se le due rotture non sono troppo vicine nello spazio e nel tempo, i meccanismi di riparazione riportano il cromosoma alla sua condizione normale. Se quelle lesioni sono invece vicine nello spazio e nel tempo, si possono verificare aberrazioni per vari tipi di riparazione errata. L'illustrazione è basata sulle ricerche di Malcom L. Randolph e J. G. Brewen dello Oak Ridge National Laboratory.**

interni da queste tre sorgenti, per una persona che vive al livello del mare, è di circa 0,8 millisievert (80 millirem) all'anno. Chi vive ad altitudini superiori, dove i raggi cosmici sono più intensi, o in regioni dove il contenuto in radio della crosta terrestre è piuttosto elevato, può ricevere una dose anche doppia.

Il contributo maggiore, fra le sorgenti artificiali, è imputabile alla diagnosi medica, la cui stima pare si avvicini alla dose ricevuta dal fondo naturale. (I contributi dovuti ai diversi tipi di indagini mediche e odontoiatriche differiscono fra loro largamente.) Fra le altre sorgenti prodotte dall'uomo vi sono i minerali radioattivi presenti in natura in taluni tipi di rocce frantumate, in materiali da costruzione e nei fertilizzanti contenenti fosfati; componenti che emettono radiazioni nei televisori, rivelatori di fumo e altri prodotti per il consumo; ricaduta da armi atomiche e perdite dai reattori nucleari.

I raggi X sono stati usati nella diagnosi e nel trattamento delle malattie già pochi mesi dopo la loro scoperta a opera di W. K. Röntgen nel 1895, e quasi subito sono stati rilevati effetti collaterali dannosi della radiazione. Il primo di tali effetti, che compariva spesso nel giro di poche ore o di pochi giorni dopo l'esposizione, era costituito da danni passeggeri come arrossamento della pelle e formazione di vesciche. Nel 1902 ci si rese conto che un effetto ritardato della radiazione poteva essere l'insorgere di tumori. Il primo caso noto di questo genere fu il tumore che colpì la mano di un radiologo, là dove questi aveva subito a lungo danni alla pelle. Nei decenni successivi, prima che venissero adottate misure di sicurezza adeguate, furono rilevate decine di tumori analoghi fra i primi addetti agli esami radiologici. Poiché questi tumori erano preceduti, in modo caratteristico, da una lunga storia di progressivi danneggiamenti alla pelle indotti dalla radiazione, si assumeva in generale che la radiazione determinasse la formazione di tumori solo se la dose era abbastanza elevata da causare forti danni. Si trattava di una forma del concetto di soglia.

La possibilità che anche piccole dosi di radiazione potessero dar luogo a tumori non fu presa in considerazione seriamente se non negli anni cinquanta, quando E. B. Lewis del California Institute of Technology dedusse (in base all'aumentata frequenza di leucemia nei sopravvissuti alla bomba atomica in Giappone, nei radiologi e in alcuni gruppi di pazienti trattati con raggi X per condizioni diverse dai tumori) che l'incidenza dei tumori poteva aumentare come una funzione lineare, priva del valore di soglia, della dose. Una tale relazione fra dose ed effetto sembrava coerente con l'ipotesi che la leucemia potesse essere l'esito di una mutazione, indotta da radiazione, in una singola cellula ematica. In base a questa idea Lewis ipotizzò che, fra tutti i casi di leucemia infantile, dal 10 al 20 per cento potessero essere attribuibili alla radiazione naturale di fondo.

L'interpretazione di Lewis, che implica che nessuna dose di radiazione è priva di qualche rischio cancerogeno, produsse preoccupazione e un dibattito che non è ancora concluso.

Anche se nel fissare limiti di esposizione alla radiazione ionizzante non si ipotizza più l'esistenza di un limite di soglia per gli effetti cancerogeni, si continua a indagare sul livello di rischio che si può attribuire alle radiazioni di basso livello.

Da molto tempo si ipotizza che gli effetti ereditari della radiazione ionizzante (cioè gli effetti che vengono trasmessi ai discendenti) non abbiano soglia e, pertanto, rappresentino un rischio associato alla



radiazione di basso livello. D'altro lato, taluni effetti non ereditari sono stati documentati clinicamente solo a livelli elevati; in questo caso sembrerebbe consistente l'ipotesi di un livello di soglia. In questa categoria rientrano l'opacizzazione del cristallino dell'occhio, la diminuzione della fecondità e l'abbassamento del numero dei globuli bianchi del sangue. Analogamente, disturbi indotti da radiazione nello sviluppo degli embrioni sono l'esito evidente di dosi abbastanza elevate da uccidere molte cellule nei tessuti colpiti e si considera pertanto improbabile che possano essere causati dall'esposizione ai bassi livelli caratteristici della radiazione naturale di fondo.

Gli effetti cancerogeni e quelli ereditari costituiscono pertanto i principali rischi connessi con piccole dosi di radiazione. Per valutare i rischi delle irradiazioni di basso livello bisogna analizzare la natura di questi effetti e quale sia la loro relazione con l'esposizione alla radiazione di un singolo individuo. Un'analisi di questo tipo richiede un esame un po' più particolareggiato delle interazioni fra la radiazione ionizzante e le cellule viventi.

Allorché la radiazione ionizzante penetra in un tessuto vivente cede la sua energia attraverso una serie di collisioni casuali e di interazioni con gli atomi e le molecole che si trovano sul suo percorso. Le collisioni e le interazioni producono ioni e radicali chimici reattivi, che scindono i legami chimici e determinano altre variazioni nelle molecole che si trovano loro vicine. La distribuzione di eventi di ionizzazione lungo il percorso del raggio o della particella in arrivo dipende dall'energia, dalla massa e dalla carica elettrica della radiazione e in minor misura dalla densità del tessuto assorbente. In generale i raggi X e i raggi gamma (che sono elettricamente neutri) sono caratterizzati da un basso tasso di trasferimento lineare di energia, cioè producono ioni qua e là lungo il loro percorso e penetrano in profondità nel tessuto. Le particelle cariche sono caratterizzate da un trasferimento lineare di energia superiore e da una minore penetrazione. La capacità di provocare danno è in rapporto con la densità della deposizione di energia. Radiazioni a elevato trasferimento lineare di energia (come i protoni e le particelle alfa) in genere inducono più danni delle radiazioni a basso trasferimento lineare di energia (come i raggi X e i raggi gamma).

Un danno cellulare è il risultato di cambiamenti molecolari provocati da ioni e radicali liberi indotti dalla radiazione. Fra i molti tipi di molecole influenzate dalla radiazione ionizzante la più critica è il DNA, a causa della limitata ridondanza di informazione genetica che è codificata in esso. La quantità totale di energia depositata da una dose particolarmente letale di raggi X che colpisce tutto il corpo (da 3 a 5 sievert circa) provoca centinaia di rotture nelle molecole di DNA in ogni cellula dell'organismo.

Lungo il percorso di una radiazione a basso trasferimento lineare di energia, è probabile che il danno al DNA sia costituito da lesioni abbastanza semplici come rotture di un singolo filamento nel doppio filamento della doppia elica del DNA. Nel caso di radiazioni a più elevato trasferimento lineare di energia è probabile che si verifichino lesioni più gravi. Dato che le lesioni di tipo semplice possono essere riparate in misura notevole da enzimi presenti nelle cellule, il danno che in ultima analisi è stato provocato al DNA in un caso particolare può dipendere tanto dall'efficacia dei processi di riparazione quanto dalla natura della lesione iniziale. Gli effetti di una lesione non riparata o riparata malamente in una singola molecola di DNA possono essere amplificati molte volte allorché il DNA viene trascritto e tradotto; alla fine quegli effetti possono essere trasmessi a un numero enorme di cellule figlie. Fosse solo per questo motivo, il danno a cromosomi e geni risulta molto più evidente di ogni altro tipo di danno causato dalla irradiazione di basso livello.

Le aberrazioni cromosomiche costituiscono uno degli effetti della radiazione studiati più a fondo. Fra tali aberrazioni vi sono variazioni nel numero e nella struttura dei cromosomi, che derivano dalla rottura e dalla ridisposizione di fibre cromosomiche e dall'interferenza con il normale processo di segregazione dei cromosomi allorché la cellula si divide.

Con processi di riparazione metabolica una cellula può far ricongiungere le estremità di un cromosoma rotto. Quando però si verificano due rotture abbastanza vicine l'una all'altra nello spazio e nel tempo, può accadere che le estremità di una rottura vengano erroneamente congiunte alle estremità di un'altra, dando luogo così a traslocazioni, inversioni, anelli e altri tipi di ridisposizioni strutturali. La frequenza di aberrazioni cromosomiche di questo tipo aumenta come funzione lineare, senza limite di soglia, della dose di radiazione nell'intervallo da basso a intermedio. Quando si riporta in grafico il numero dei danni cromosomici, la pendenza della curva è maggiore per la radiazione a elevato trasferimento lineare di energia e minore per la radiazione a basso trasferimento di energia. All'aumentare della dose e dell'intensità di dose la pendenza per la radiazione a basso trasferimento aumenta, ma resta sostanzialmen-

**Viene mostrato, sulla base di varie stime, il contributo della esposizione a radiazioni al numero totale di tumori letali nella popolazione degli USA. La dose annuale da ciascuna sorgente di radiazione è espressa in sievert-persona. Il numero totale di tumori letali è inferiore a 10 000 all'anno, cioè rappresenta circa il 2,5 per cento della mortalità annua negli USA a seguito di tumori dipendenti da cause sia note sia ignote.**

te immutata per la radiazione a elevato trasferimento.

In base a queste relazioni fra dose ed effetto si può dedurre che la probabilità che una aberrazione cromosomica sia causata da una radiazione a basso trasferimento di energia è bassa, a meno che due o più percorsi di radiazione passino attraverso una data regione, nel nucleo della cellula, in successione abbastanza ravvicinata perché le lesioni cromosomiche provocate lungo un percorso possano interagire con quelle provocate lungo un altro percorso. Le lesioni risultanti dal passaggio di una singola radiazione a elevato trasferimento hanno, invece, una alta probabilità di interagire le une con le altre, indipendentemente dalle lesioni che vengono prodotte da un altro passaggio di radiazioni.

La frequenza di aberrazioni cromosomiche in linfociti (un tipo di globuli bianchi del sangue) prelevati nel flusso sanguigno e conservati in un mezzo di coltura è sensibilmente maggiore nelle persone che lavorano a contatto con la radiazione e in altre persone esposte a radiazione. La frequenza combinata di tali aberrazioni, stimata immediatamente dopo irradiazione in vitro, si avvicina a 0,1 per cellula per sievert; da questo valore si può calcolare che la dose necessaria per raddoppiare la loro frequenza sis vicina a 0,05 sievert, cioè circa 60 volte la dose ricevuta normalmente ogni anno dalla radiazione naturale di fondo. La frequenza delle aberrazioni di questo tipo è aumentata anche da taluni virus, prodotti chimici e farmaci. Nella maggior parte dei casi queste aberrazioni sono transitorie, in quanto interferiscono con il processo di divisione delle cellule colpite; tali cellule non riescono a riprodursi e a poco a poco scompaiono.

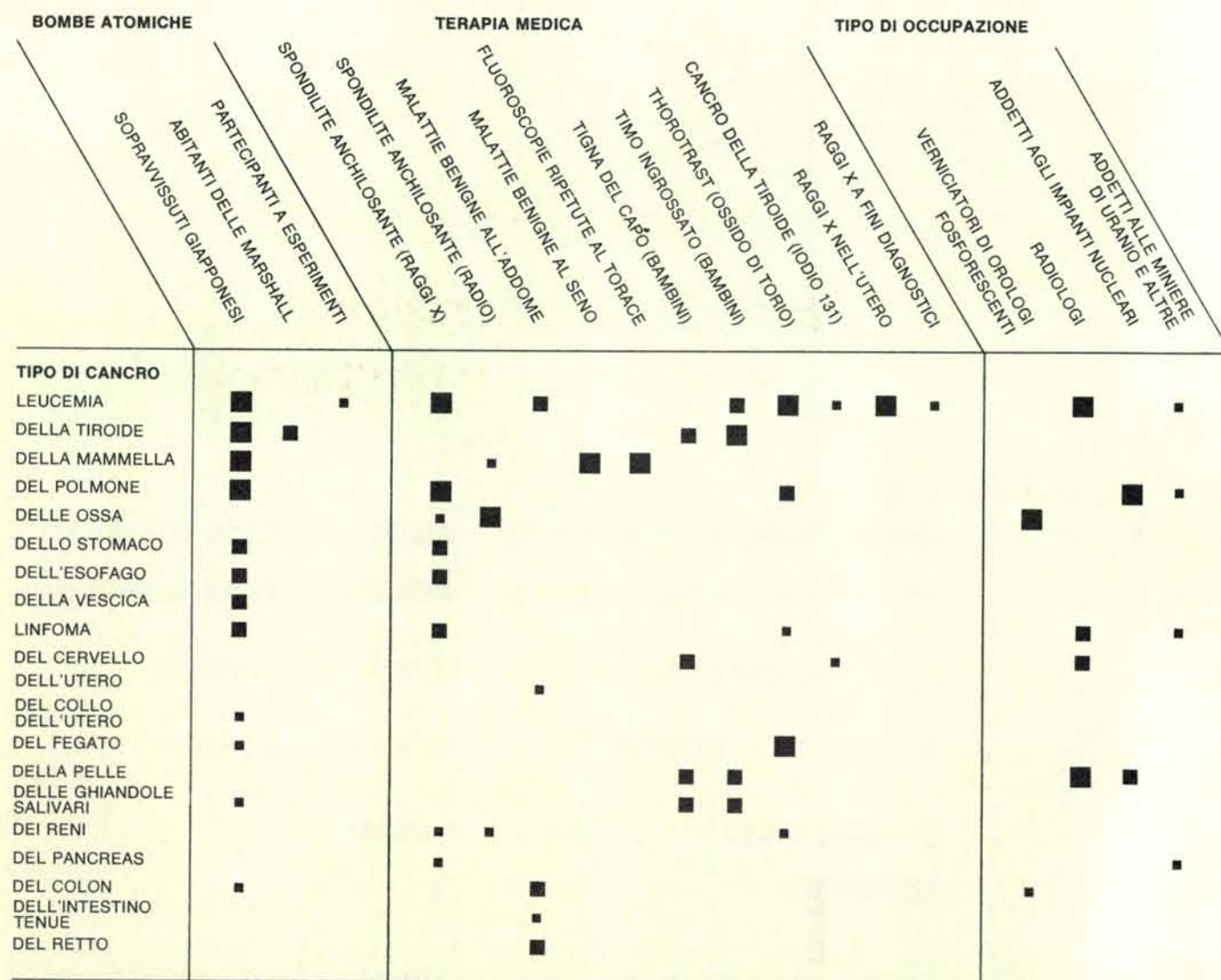
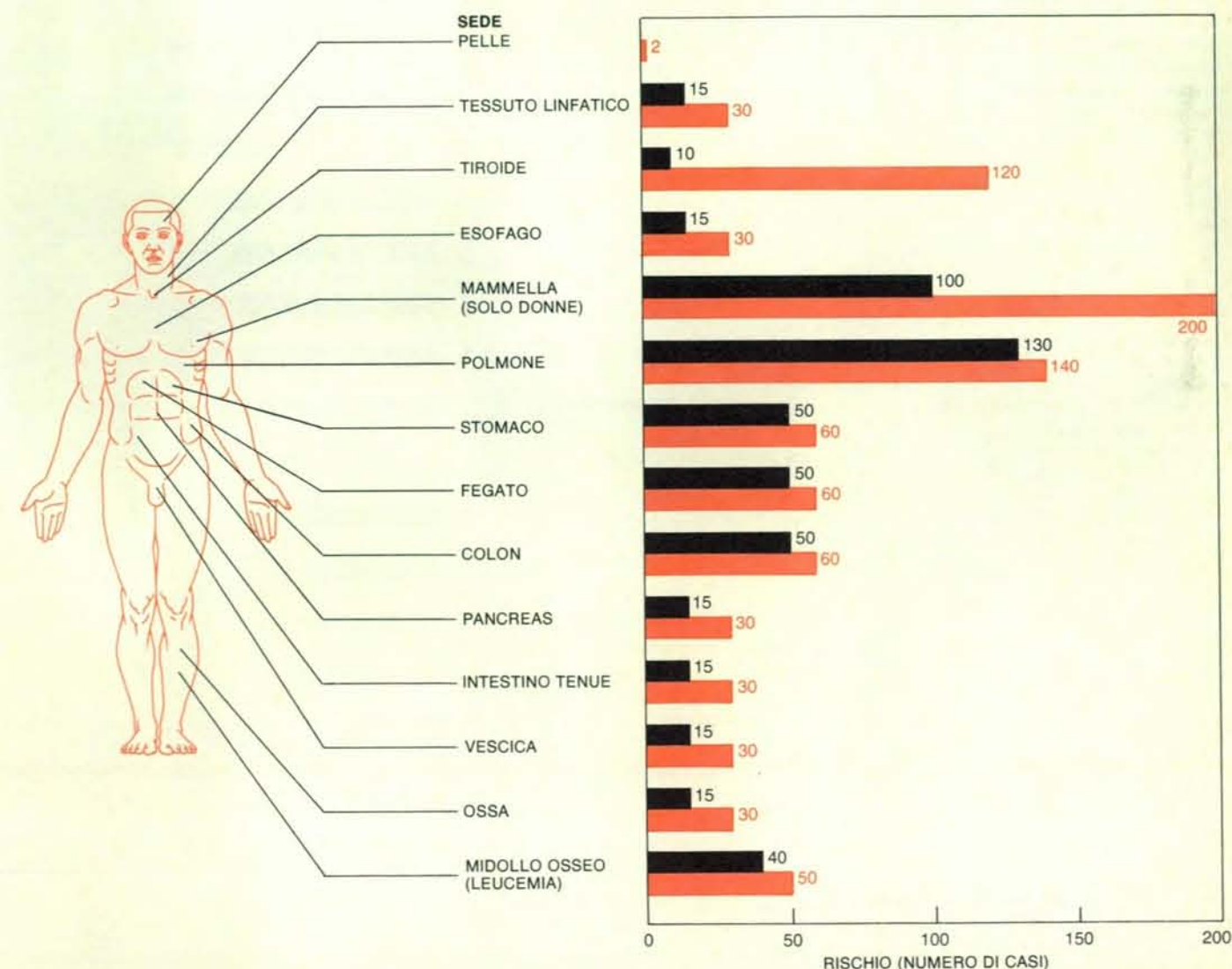
L'azione mutagena della radiazione sui geni è stata ampiamente studiata in un gran numero di organismi diversi, dopo la sua scoperta nel 1927 a opera di H. J. Muller. Fino a oggi non sono stati dimostrati effetti ereditari della radiazione in popolazioni umane, ma l'abbondanza di informazioni pertinenti ottenute da altre specie animali ci fornisce una base per prevedere l'insorgere di tali effetti anche nell'uomo. I dati più pertinenti sono quelli ottenuti dagli esperimenti con i topi.

In questi esperimenti la relazione tra dose ed effetto per l'induzione di mutazioni in spermatogoni e oociti (spermatozoi e uova in maturazione) è molto simile a quella per le aberrazioni cromosomiche. Con la radiazione a elevato trasferimento di energia, la frequenza delle mutazioni cresce rapidamente proporzionalmente con la dose, ma è relativamente indipendente dall'intensità di dose. (La dose è la quantità totale di energia di radiazione depositata in un tessuto o in una cellula indipendentemente dal tempo; l'intensità di dose è la quantità di energia di radiazione per unità di tempo.) Con la radiazione a basso trasferimento di energia, la frequenza delle mutazioni aumenta meno rapidamente in funzione della dose, ma è fortemente dipendente dall'intensità di dose.

Quando si studiano questi problemi l'attenzione si rivolge agli effetti genetici dei bombardamenti atomici e della radiazione naturale di fondo. Fra i figli dei sopravvissuti ai due bombardamenti atomici non è comparso alcun aumento apprezzabile nelle anomalie genetiche. Il fatto di non essere riusciti a rilevare alcun aumento non è incompatibile con le osservazioni contrastanti effettuate nelle ricerche sui topi, viste le piccole dimensioni del campione (78 000 bambini) e la bassa dose media (0,5 sievert) alle gonadi dei genitori.

Le informazioni attualmente disponibili fanno pensare che la dose necessaria per raddoppiare il tasso di mutazione nell'uomo sia compresa fra 0,2 e 2,5 sievert. Sulla base di questi valori sono state effettuate valutazioni del contributo della radiazione naturale di fondo sulla frequenza di vari tipi di malattie, determinate geneticamente, nella popolazione. Queste valutazioni sono molto incerte, poiché è difficile mettere in rapporto il numero delle malattie genetiche in una popolazione con il solo tasso di mutazione, ma esse comportano, tuttavia, che solo un piccolo numero (dallo 0,1 al 2 per cento) di tutte le malattie determinate geneticamente sia imputabile alla radiazione naturale di fondo.

La radiazione aumenta la frequenza della maggior parte dei tumori, ma non di tutti. L'incremento dipende dalle condizioni di esposizione e da fattori che influenzano la predisposizione della popolazione esposta. Una seria difficoltà è dovuta al fatto che tumori indotti da radiazione non presentano caratteristiche distintive in base alle quali possano essere identificati come tali. Inoltre, non compaiono se non anni o decenni dopo l'irradiazione. Pertanto si può dedurre una relazione causale fra tumore e precedente irradiazione solo sulla base di opportune prove epidemiologiche. Le tre fonti migliori per prove di questo genere sono gli studi condotti sui sopravvissuti ai bombardamenti atomici, su pazienti esposti a radiazione a fini medici e su lavoratori esposti alla radiazione negli ambienti di lavoro.

Poiché la frequenza di tumori di qualunque tipo è relativamente bassa, l'analisi della relazione matematica fra dose e incidenza richiede lo studio di grandi popolazioni. Inoltre, dato il gran numero di soggetti e il tempo che intercorre fra l'irradiazione e l'insorgere del tumore, studi di questo genere di solito sono complicati da incertezze relative all'entità della dose di radiazione che ha colpito i soggetti e alla possibile influenza di rischi di natura diversa dalle radiazioni. Ciononostante, vi è una notevole coeren-

**È qui riportata l'incidenza di neoplasie associate a radiazioni per persone esposte a radiazione in tre circostanze diverse; bombardamenti atomici nel Giappone, radioterapie ed esposizione professionale. Una forte associazione fra un certo tipo di tumore e un certo tipo di esposizione è indicata dai quadrati più grandi, una associazione significativa dai quadrati di media dimensione e una associazione possibile, ma non confermata, dai quadrati più piccoli. I tre gruppi presi in considerazione sono quelli per cui l'esposizione a radiazione è ben documentata.**

**È qui indicato il rischio nell'arco di tutta la vita per vari tipi di tumore dovuti a radiazioni di basso livello, sulla base di stime di vari ricercatori. Il rischio si riferisce a 10 000 sievert-persona, cioè all'equivalente di un sievert di radiazione per 10 000 persone nell'arco di tutta la vita. Le cifre indicate rappresentano le stime massime per tumori letali *(in nero)* e il numero di casi di tumore *(in colore)*. In ogni esempio le stime coprono in realtà un intervallo in cui il valore minimo è molto inferiore (per esempio, 15 casi e 15 decessi per il tumore al midollo osseo).**

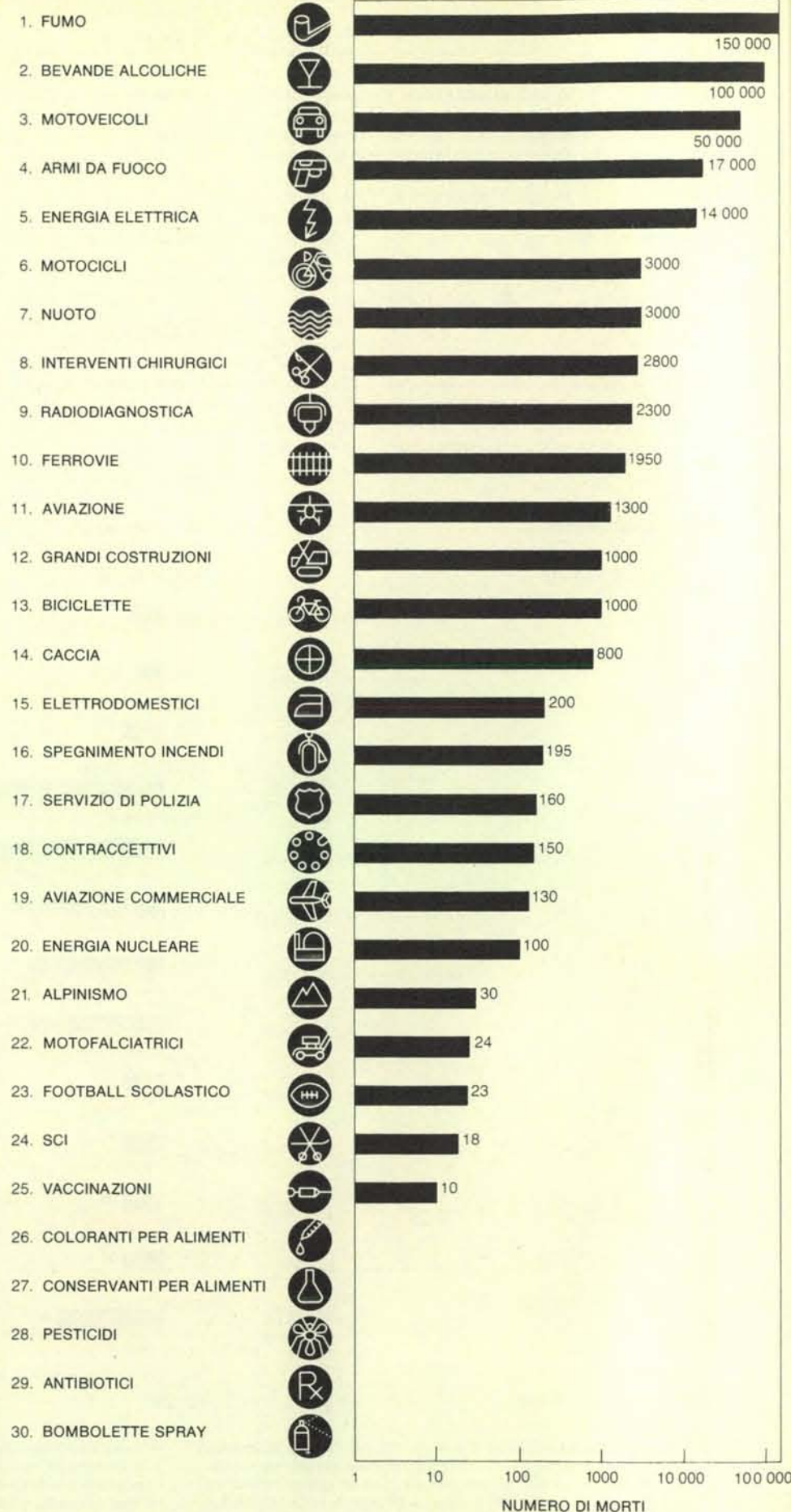
za fra differenti popolazioni irradiate per quel che riguarda la misura con la quale l'insorgenza di tumori varia in rapporto con la dose stimata di radiazione. Bisogna dire però che i dati si riferiscono prevalentemente a dosi di radiazioni piuttosto elevati (da 0,5 a 2 sievert), cosicché i rischi cancerogeni della radiazione di basso livello possono essere stimati soltanto con una estrapolazione basata su ipotesi approssimative in merito alla relazione fra dose e incidenza.

Disponiamo di informazioni in merito a questa relazione per la leucemia e per il tumore della mammella nella donna più che per ogni altro tipo di tumore umano. I dati sono relativi a diverse popolazioni e coprono una gamma notevole di dosi. Nel caso della leucemia, l'incidenza di tutte le forme della malattia, fatta eccezione per la forma linfatica cronica, cresce con la dose. Nei primi 25 anni dopo l'irradiazione l'eccesso cumulativo totale di tutte le forme combinate è di circa un caso all'anno su 10 000 persone a rischio per sievert di radiazione al midollo osseo.

La coerenza delle cifre per il tumore della mammella da gruppo a gruppo è notevole, considerate le differenti durate di esposizione fra le varie popolazioni. L'irradiazione è stata quasi istantanea per persone esposte a bombardamenti atomici, è stata ricevuta in più esposizioni (da 1 a 4) in giornate successive da parte di donne sottoposte a radioterapia ed è stata accumulata nell'arco di molti mesi (con incrementi di 0,01-0,05 sievert) da donne sottoposte più volte a esami fluoroscopici del torace nel corso di cure per tubercolosi polmonare. (Il trattamento comportava il ricorso al pneumotorace; gli esami fluoroscopici venivano effettuati per essere sicuri che il polmone non avesse ricominciato a espandersi.) Il fatto che l'aumento dei casi di tumore per dose unitaria non sia sensibilmente minore nel terzo gruppo rispetto agli altri fa pensare che esposizioni di piccola entità, successive e ampiamente intervallate l'una dall'altra, siano completamente additive nel loro effetto complessivo sul tessuto della mammella. L'assenza di riparazioni rilevabili fra queste piccole esposizioni ampiamente intervallate è in buon accordo con una relazione lineare, senza limite di soglia, fra dose e incidenza.

Disponiamo di prove ulteriori in merito a effetti cancerogeni da basse dosi grazie agli studi sul tumore della tiroide insorto in seguito a radioterapia e sull'associazio-

| LEGA DELLE DONNE ELETTRICI | STUDENTI UNIVERSITARI | MEMBRI DI ASSOCIAZIONI COMMERCIALI E PROFESSIONALI |
|---|---|---|
| ENERGIA NUCLEARE | ENERGIA NUCLEARE | ARMI DA FUOCO |
| MOTOVEICOLI | ARMI DA FUOCO | MOTOCICLI |
| ARMI DA FUOCO | FUMO | MOTOVEICOLI |
| FUMO | PESTICIDI | FUMO |
| MOTOCICLI | MOTOVEICOLI | BEVANDE ALCOLICHE |
| BEVANDE ALCOLICHE | MOTOCICLI | SPEGNIMENTO INCENDI |
| AVIAZIONE | BEVANDE ALCOLICHE | SERVIZIO DI POLIZIA |
| SERVIZIO DI POLIZIA | SERVIZIO DI POLIZIA | ENERGIA NUCLEARE |
| PESTICIDI | CONTRACCETTIVI | INTERVENTI CHIRURGICI |
| INTERVENTI CHIRURGICI | SPEGNIMENTO INCENDI | CACCIA |
| SPEGNIMENTO INCENDI | INTERVENTI CHIRURGICI | AVIAZIONE |
| GRANDI COSTRUZIONI | CONSERVANTI PER ALIMENTI | ALPINISMO |
| CACCIA | BOMBOLETTE SPRAY | GRANDI COSTRUZIONI |
| BOMBOLETTE SPRAY | GRANDI COSTRUZIONI | BICICLETTE |
| ALPINISMO | AVIAZIONE | PESTICIDI |
| BICICLETTE | AVIAZIONE COMMERCIALE | SCI |
| AVIAZIONE COMMERCIALE | RADIODIAGNOSTICA | NUOTO |
| ENERGIA ELETTRICA | CACCIA | AVIAZIONE COMMERCIALE |
| NUOTO | ENERGIA ELETTRICA | ENERGIA ELETTRICA |
| CONTRACCETTIVI | COLORANTI PER ALIMENTI | FERROVIE |
| SCI | ANTIBIOTICI | FOOTBALL SCOLASTICO |
| RADIODIAGNOSTICA | ALPINISMO | CONTRACCETTIVI |
| FOOTBALL SCOLASTICO | FERROVIE | BOMBOLETTE SPRAY |
| FERROVIE | BICICLETTE | RADIODIAGNOSTICA |
| CONSERVANTI PER ALIMENTI | SCI | MOTOFALCIATRICI |
| COLORANTI PER ALIMENTI | FOOTBALL SCOLASTICO | ANTIBIOTICI |
| MOTOFALCIATRICI | ELETTRODOMESTICI | ELETTRODOMESTICI |
| ANTIBIOTICI | MOTOFALCIATRICI | CONSERVANTI PER ALIMENTI |
| ELETTRODOMESTICI | VACCINAZIONI | VACCINAZIONI |
| VACCINAZIONI | NUOTO | COLORANTI PER ALIMENTI |

**La percezione del rischio varia notevolmente in gruppi diversi di persone. I diagrammi a destra riflettono il risultato di un'indagine condotta dalla Decision Research in cui si chiedeva a persone di tre gruppi di classificare le trenta sorgenti di rischio elencate nella prima colonna. In questa colonna le fonti sono ordinate secondo il loro effettivo contributo annuo al numero dei decessi negli USA, come è stato determinato da valutazioni attuariali. Questo contributo è riportato nella seconda colonna, mentre nelle altre colonne sono registrate le classifiche scelte dai tre gruppi di intervistati.**

| SITUAZIONE | CAUSA DI MORTE |
|---|---|
| VIAGGIARE PER 1000 CHILOMETRI IN AEREO | INCIDENTE |
| ATTRAVERSARE L'OCEANO IN AEREO | TUMORE DA RAGGI COSMICI |
| VIAGGIARE PER 90 CHILOMETRI IN AUTOMOBILE | INCIDENTE |
| SOGGIORNO DI DUE MESI A DENVER | TUMORE DA RAGGI COSMICI |
| VIVERE PER DUE MESI IN UN EDIFICIO DI PIETRA | TUMORE DA RADIOATTIVITÀ |
| LAVORARE PER 10 GIORNI IN UN'INDUSTRIA | INCIDENTE |
| LAVORARE PER TRE ORE IN UNA MINIERA DI CARBONE | INCIDENTE |
| FUMARE DA 1 A 3 SIGARETTE | TUMORE, MALATTIE CARDIACHE E POLMONARI |
| SCALARE UNA MONTAGNA PER 1,5 MINUTI | INCIDENTE |
| VENTI MINUTI ALL'ETÀ DI 60 ANNI | QUALSIASI |

**Il rischio da radiazione è qui confrontato con quello dovuto ad altre situazioni o attività. In ciascun caso il rischio è di una possibilità su un milione di morire a seguito dell'esposizione. Lo schema è basato sui lavori di E. E. Pochin in Inghilterra e Richard Wilson della Harvard University.**

ne fra irradiazione prenatale e tumori dell'infanzia. In vari studi indipendenti si è trovato che l'incidenza dei tumori della tiroide aumenta dopo un trattamento locale alla testa e al collo con raggi X nel corso dell'infanzia. Due di questi studi hanno mostrato che l'aumento dei tumori è evidente già nel caso di dosi alla tiroide comprese fra 0,06 e 0,2 sievert.

La associazione fra irradiazione prenatale e tumori infantili può non essere causale, ma non sono state ancora dimostrate altre cause. Presi così come appaiono, i dati comportano che la somministrazione di raggi X a scopo diagnostico durante lo sviluppo intrauterino possa aumentare anche del 50 per cento il rischio di sviluppo di un tumore prima della pubertà. Poiché la dose media in questi casi varia solamente da 0,01 a 0,05 sievert, il numero dei tumori in eccesso fa pensare che l'embrione e il feto siano relativamente suscettibili agli effetti cancerogeni della radiazione.

Nelle popolazioni irradiate si è osservato un aumento in frequenza di molti altri tipi di tumore, ma in genere i dati relativi al rapporto fra dose e incidenza per questi tumori sono limitati. Ciononostante, gli stessi indicano che gli effetti cancerogeni di una data dose solitamente sono maggiori nel caso di radiazione ad alto trasferimento di energia che non nel caso di radiazione a basso trasferimento. Indicano inoltre che l'aumento nel numero dei tumori per dose unitaria varia sensibilmente a seconda della localizzazione del tumore.

Gruppi di ricercatori che studiano gli effetti biologici della radiazione ionizzante per conto delle Nazioni Unite e della US National Academy of Sciences hanno analizzato i dati disponibili e hanno così stimato che l'eccesso totale di tumori di tutti i tipi sia di 4-18 (2-10 dei quali fatali) all'anno su 10 000 persone a rischio per sievert, a partire da circa 2-10 anni dopo l'irradiazione di tutto il corpo; tale incremento si conserverebbe quindi per tutto il resto della vita della popolazione a rischio. Queste cifre corrispondono a un incremento cumulativo, lungo tutto l'arco della vita, di 140-1000 tumori (da 70 a 500 dei quali fatali) su 10 000 persone a rischio per sievert.

Queste stime sono state criticate perché sottovaluterebbero il rischio effettivo, forse anche di un fattore 10, poiché sembra trascurino certi studi di popolazioni esposte a basse dosi di radiazione in cui sono stati osservati più tumori di quelli che si sarebbero potuti prevedere in base alle stime. Questi studi riguardano fra l'altro addetti dell'impianto nucleare di Hanford (Washington), addetti al settore nucleare dei cantieri navali di Portsmouth, nel New Hampshire, personale militare esposto a esperimenti di armi atomiche nel Nevada, bambini in zone sottovento rispetto a un sito di esperimenti nucleari nello Utah e pazienti trattati con raggi X a fini diagnostici.

Qui sarà sufficiente dire che nessuno di questi studi riesce a porre in correlazione in modo convincente i tumori in questione con stime affidabili della dose di radiazione e nessuno di essi tiene adeguatamente conto della possibile influenza della esposizione a fumo di sigarette, farmaci, prodotti chimici industriali o altre variabili che possono complicare la situazione. Secondo l'opinione della maggior parte degli esperti questi studi, pertanto, non sono sufficientemente convincenti da costringerci ad aumentare le stime attuali dei rischi da radiazione. Al contrario, secondo molti esperti le stime attuali esagerano il rischio mancando di tener conto opportunamente delle riparazioni dei danni a basse dosi e a basse intensità di dose, che sono state invece osservate nei tumori indotti in animali da esperimento.

Nel caso di radiazioni a basso trasferimento lineare di energia esistono buoni motivi di natura radiobiologica per aspettarsi che il rischio per dose unitaria vada diminuendo al diminuire della dose e dell'intensità di dose. In questi casi, quindi, l'estrapolazione mediante un modello lineare senza limite di soglia tenderebbe a sopravvalutare il rischio di esposizione a bassi livelli di radiazioni di questo tipo; sarebbero più corrette, pertanto, stime un poco inferiori. Nel caso di radiazione a elevato trasferimento lineare di energia, invece, i dati implicano che possa essere vero il contrario, dal momento che gli effetti cancerogeni di questa radiazione sembrano aumentare al diminuire della dose e dell'intensità di dose.

Se conoscessimo meglio i meccanismi delle diverse forme di tumore indotto da radiazione, tutti gli ipotetici modelli della relazione fra dose e incidenza risulterebbero probabilmente una grossolana ipersemplificazione. Le ricerche su animali da esperimento nel caso di dosi intermedie ed elevate indicano che queste dosi possono contribuire al rischio di tumori diversi mediante differenti effetti sull'equilibrio endocrino, sull'immunità, sul ricambio delle cellule e su altre funzioni fisiologiche, con relazioni fra dose e incidenza che possono variare corrispondentemente. Dati relativi ai tumori dello scheletro, alla leucemia e al tumore della mammella provano che la relazione fra dose e incidenza può variare in maniera analoga anche nell'uomo. Riportando in grafico l'incidenza di questi tumori si ottengono curve simili per la leucemia e il tumore della mammella, ma una curva nettamente diversa per i tumori dello scheletro.

Il numero dei tumori che possono essere attribuiti alla radiazione di basso livello negli adulti è piccolo rispetto al numero di quelli che possono essere imputati ad altre cause; tuttavia qualunque studio epidemiologico che cerchi di verificare le stime di rischio per dosi comprese fra 1 e 50 millisievert dovrebbe passare in rassegna centinaia di migliaia di soggetti. In futuro, pertanto, i tentativi di migliorare le stime del rischio per l'uomo non dovrebbero basarsi esclusivamente su osservazioni empiriche, ma dovrebbero comprendere anche ricerche sui meccanismi degli effetti cancerogeni in animali da laboratorio, in cellule in coltura e in altri sistemi sperimentali.

In base alle attuali stime del rischio e assumendo che il rischio vari come funzione lineare, senza limite di soglia, della dose fino al livello della radiazione naturale di fondo, si può dedurre che, di tutti i tumori di una popolazione, solo una piccola percentuale (dall'1 al 3 per cento) è imputabile alla radiazione naturale di fondo. Questa deduzione è in accordo con il confronto empirico fra incidenza di tumori in aree geografiche che differiscono fra loro per il livello della radiazione di fondo. Le stime comportano inoltre che una percentuale sensibile (forse anche del 20 per cento o più) dei tumori del polmone nei non fumatori può essere imputata alla inalazione di isotopi radioattivi presenti naturalmente nell'aria.

Il numero dei tumori imputabili a irradiazione sul posto di lavoro nella maggior parte dei gruppi di lavoratori a contatto con radiazioni rappresenta un incremento inferiore all'1 per cento rispetto all'incidenza naturale. Pertanto il rischio professionale medio per i lavoratori a contatto con radiazioni, espresso in termini di diminuzione della speranza di vita per effetti cancerogeni ed ereditari combinati, è paragonabile al rischio che si corre in occupazioni di solito classificate come sicure.

Se si pensa che nessuna quantità di radiazione è del tutto priva di effetti, si deve presumere l'esistenza di qualche rischio associato con tutte le attività che comportano irradiazioni di basso livello, indipendentemente dall'esiguità della dose. Alcuni di questi rischi, a confronto con altri rischi della vita quotidiana, sembrano trascurabili *(si veda l'illustrazione alla pagina precedente)*. Inoltre, nessun rischio deve essere considerato accettabile se può essere facilmente evitato o se non è accompagnato da un beneficio corrispondente. In generale l'accettabilità di un rischio tende a diminuire in funzione della involontarietà, in particolare se è imposto a coloro che non partecipano equamente alla divisione dei benefici relativi. La valutazione dei rischi e dei benefici richiede un giudizio personale di valore, che può variare ampiamente da individuo a individuo.

La valutazione dei rischi sanitari derivanti dalla radiazione di basso livello è un problema scientifico, ma l'accettabilità di questi rischi dipende chiaramente dalla grandezza percepita, oltre che dalla grandezza reale, dei rischi e dei benefici associati. Pertanto la decisione sull'accettabilità dei rischi da radiazione è un processo in cui deve essere rappresentata la società tutta. La possibilità di decisioni corrette dipenderà dalla comprensione da parte del pubblico dei problemi in gioco, oltre che dai dati forniti dalla scienza.

È degno di nota, inoltre, che i problemi che si incontrano nel valutare i rischi delle radiazioni di basso livello siano paralleli a problemi analoghi che si incontrano nella valutazione dei rischi connessi con l'esposizione a bassi livelli di sostanze chimiche tossiche. Questi ultimi problemi sono anche più complessi: coinvolgono non soltanto variazioni poco definite fra gli individui e fra le specie per quel che riguarda l'assunzione, la distribuzione, il metabolismo, la disintossicazione e l'escrezione di sostanze, ma anche altre fonti di incertezza. Ciononostante, le basi scientifiche per valutare i rischi delle radiazioni e delle sostanze chimiche sono abbastanza simili perché i concetti e i principi formulati nel tentativo di comprendere e proteggere contro gli effetti biologici della radiazione possano essere utili nel formulare strategie per la protezione contro i rischi sanitari di sostanze chimiche tossiche e di altri agenti ambientali potenzialmente dannosi.

# Come un virus animale entra ed esce dalla cellula ospite

*Gli esperimenti realizzati con il virus Semliki Forest mostrano come la cellula ospite sia costretta a fabbricare nuove particelle virali, giungendo anche a cedere un frammento della propria membrana esterna*

di Kai Simons, Henrik Garoff e Ari Helenius

Per una pianta o un animale, la cellula è l'unità fondamentale di struttura e di funzione. Per un virus, essa è semplicemente un tramite per produrre nuove particelle virali. Una particella virale consiste di uno o più filamenti di acido nucleico (DNA o RNA) racchiusi in un «guscio» di natura proteica, detto capside. In molti virus il capside stesso è avvolto in una membrana contenente molecole proteiche, esattamente come la membrana esterna di una cellula animale. L'acido nucleico presente nella particella contiene l'informazione genetica necessaria per la duplicazione del virus. La particella manca, però, dell'apparato capace di realizzare la sua struttura definitiva partendo da tale informazione e utilizza quindi quello della cellula ospite. Il virus penetra in essa e ne espropria la macchina biosintetica, per poter fabbricare le proprie parti. Una volta prodotte in numero sufficiente, esse vengono assemblate per costituire nuove particelle virali, che abbandonano poi la cellula per andare a infettarne altre.

La natura dell'infezione virale, nelle sue linee generali, è nota da vari anni, ma gli esatti meccanismi molecolari grazie ai quali il virus può entrare in una cellula e uscirne sono ancora in fase di studio. Benché non tutto il processo infettivo sia noto nei particolari, è ora estremamente chiaro che sono pochissime le reazioni del ciclo vitale del virus specifiche del virus stesso. Per lo più, il virus sfrutta le normali reazioni metaboliche del suo ospite. Nella nostra indagine condotta sul virus Semliki Forest, un virus a RNA che infetta gli animali, abbiamo appurato che esso penetra in una cellula seguendo la stessa via di molte molecole necessarie al metabolismo.

Dapprima le proteine della membrana esterna del virus si fissano alla membrana cellulare; poi questa avvolge la particella virale, formando una vescicola, che, all'interno della cellula, si fonde con altre vescicole; infine, la particella virale va a depositarsi in un grosso vacuolo, chiamato lisosoma. (L'interno del lisosoma è acido e contiene enzimi capaci di degradare grosse molecole come le proteine e gli acidi nucleici.) Solo a questo punto avviene una reazione che è specifica del virus: la membrana esterna del virus si fonde con quella del lisosoma, liberando nel citoplasma della cellula il capside e il suo contenuto di RNA, che sfuggono così all'azione degradatrice degli enzimi lisosomici. Può quindi avere inizio il processo di duplicazione della particella virale.

La replicazione dell'RNA virale e la sintesi delle proteine del capside e della membrana avvengono quasi interamente a opera dei normali processi metabolici della cellula. Il nuovo RNA e le nuove proteine del capside vengono poi assemblate nel citoplasma a formare un nuovo nucleo virale: il nucleocapside. Le proteine destinate alla membrana esterna del virus vengono sintetizzate, modificate, trasportate e inserite nella membrana cellulare, proprio come accade per le proteine della membrana cellulare stessa. Il nucleocapside neosintetizzato si sposta verso la membrana e le proteine del capside si legano alle proteine virali in essa contenuta, scartando quelle appartenenti all'ospite. A mano a mano che questa operazione procede, la membrana cellulare si incurva attorno al nucleocapside e, quando ogni molecola proteica del capside risulta legata alle proteine virali presenti nella membrana cellulare, la particella virale esce dalla cellula. Come ospite, il virus Semliki Forest è davvero bisognoso di tutto: entra nella cellula attraverso i normali processi di ingestione, ne espropria la macchina biosintetica per costruire le proprie parti e infine lascia la cellula avvolto in un frammento della sua membrana.

Il virus Semliki Forest, la cui sigla è SFV, fu isolato per la prima volta nel 1944 dai tessuti di una zanzara ed è stato chiamato con il nome della foresta pluviale nel sud dell'Uganda dove la zanzara era stata rinvenuta. I virus vengono generalmente raggruppati in tre categorie: quelli che infettano cellule animali, quelli che infettano cellule vegetali e quelli che infettano batteri. Date le differenze di composizione delle diverse cellule ospiti, i meccanismi d'ingresso e successiva liberazione dei tre tipi di virus sono alquanto diversi. Il virus Semliki Forest è un virus animale che infetta una vasta schiera di vertebrati e di invertebrati e ha una stretta affinità con i virus che causano la febbre gialla e la dengue. Può provocare l'encefalite nei topi, ma in generale non è patogeno per l'uomo. Appartiene a una classe di virus noti come «togavirus»: virus piccoli e sferici, con una struttura relativamente semplice. Il loro nome deriva dalla presenza di una membrana, una sorta di «toga», attorno al capside. Ciò li distingue da quei virus la cui superficie esterna è costituita dal capside stesso. Al centro della particella di SFV si trova una molecola di RNA a filamento singolo, una catena costituita da 12 700 nucleotidi. (Ogni nucleotide che compone la molecola di RNA consiste di una molecola di zucchero, di un gruppo fosfato e di una delle quattro basi azotate).

La relativa semplicità del virus Semliki Forest si manifesta con il fatto che ogni particella virale incorpora solo quattro tipi di molecole proteiche. Il capside è costituito da 180 molecole di una proteina, chiamata proteina del capside, o proteina *C*. Ogni molecola di proteina *C* è una catena di 267 amminoacidi e la struttura del capside è quella di un icosaedro. L'icosaedro è un solido con 20 facce, ognuna delle quali è un triangolo equilatero. Nel capside ogni faccia dell'icosaedro è suddivisa in tre triangoli più piccoli; la superficie del capside ha pertanto 60 facce. In corrispondenza di ogni angolo delle 60 facce si trova una molecola di proteina *C*, si ha quindi un totale di 180 molecole di proteina *C*.

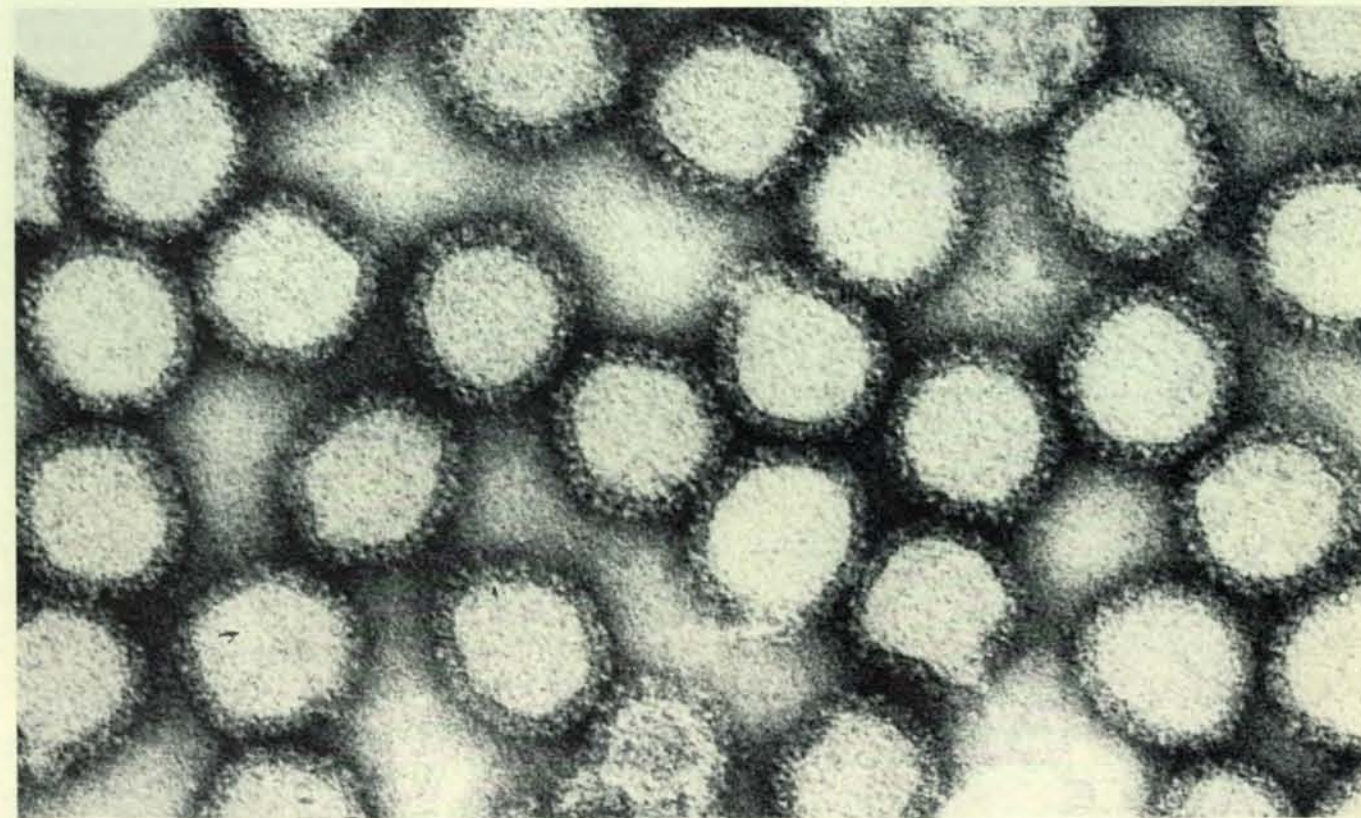
Le altre tre proteine del virus Semliki Forest formano delle sporgenze a punta lungo tutta la membrana esterna del virus. Studi biochimici hanno messo in rilievo che ogni punta è composta da una molecola di ognuna delle tre proteine, che sono state indicate come $E1$ (438 amminoacidi), $E2$ (422 amminoacidi) ed $E3$ (66 amminoacidi). Tutte e tre le proteine possiedono catene laterali di carboidrati e quindi sono, in realtà, glicoproteine.

Le glicoproteine delle punte della membrana del SFV sono attorcigliate fra loro in modo complesso. Di conseguenza, tutta la proteina $E3$ e circa il 90 per cento delle proteine $E1$ ed $E2$ sporgono dalla superficie esterna della membrana virale, formando la parte delle punte che è visibile nelle microfotografie elettroniche. I brevi segmenti di $E1$ ed $E2$ che penetrano nella membrana sono composti da due soli amminoacidi nel caso di $E1$ e da 31 amminoacidi nel caso di $E2$. Questi corti segmenti interni servono ad ancorare la punta al capside.

La sequenza degli amminoacidi nelle proteine delle punte è stata stabilita determinando quella dei nucleotidi nell'RNA virale. (Ogni amminoacido è codificato da una sequenza di tre nucleotidi.) È risultato difficile determinare la completa sequenza degli amminoacidi nelle proteine che attraversano le membrane mediante un'analisi diretta, in quanto raramente è possibile ottenerle non mescolate ad altre sostanze e in quantità sufficienti. Ma, grazie ai metodi che consentono di determinare rapidamente l'ordine dei nucleotidi in una molecola di RNA o di DNA, si è potuto negli ultimi anni, e in modo relativamente semplice, identificare indirettamente la sequenza degli amminoacidi. In collaborazione con Anne-Marie Frischauf e Hans Lehrach dello European Molecular Biology Laboratory (EMBL), ci siamo serviti di questi metodi per analizzare le proteine delle punte della membrana del virus Semliki Forest.

La nostra ricerca ha messo in rilievo che sia la proteina $E1$ sia la proteina $E2$ possiedono un segmento insolito, prevalentemente costituito da amminoacidi idrofobi. Una fila ininterrotta di questi amminoacidi raramente si trova nelle proteine idrosolubili del citoplasma, mentre è presente nelle proteine della membrana, diverse da quelle del virus Semliki Forest. I primi a determinare la sequenza degli amminoacidi in una glicoproteina della membrana, contenente un segmento idrofobo, sono stati Jere P. Segrest, Vincent T. Marchesi e collaboratori presso il National Institute of Arthritis, Metabolism and Digestive Diseases i quali si sono serviti del metodo diretto. Altre proteine di membrana (tra cui le immunoglobuline legate alle membrane, le proteine antigeniche dei trapianti e le proteine delle punte di altri virus) sono state analizzate in seguito con il metodo indiretto.

Tutte queste proteine hanno in comune con le proteine delle punte del SFV importanti proprietà strutturali: quando fanno parte della membrana cellulare, la maggior parte della loro massa sporge all'esterno della cellula e il collegamento con la membrana avviene a opera di un segmento costituito da 20-30 amminoacidi idrofobi e che attraversa la membrana; un breve troncone di pochi amminoacidi sporge all'interno nel citoplasma. La stretta somiglianza tra le proteine delle punte del virus Semliki Forest e le proteine della membrana della cellula ospite risulterà ancora più evidente quando descriveremo la sintesi delle prime.

Per quanto riguarda le molecole proteiche, il virus Semliki Forest è davvero singolare, infatti non contiene mai proteine della cellula ospite, mentre le molecole lipidiche della sua membrana hanno questa provenienza. La membrana del SFV ha la stessa conformazione di quella di una cellula animale: due strati di molecole lipidiche in cui sono inserite diverse glicoproteine. (Ogni molecola lipidica ha una «testa» idrofila e due «code» idrofobe, costituite da grosse molecole di acidi grassi. Nella membrana, le teste idrofile sono

**In questa microfotografia elettronica di Carl-Henrik von Bonsdorff dell'Università di Helsinki, le particelle del virus Semliki Forest corrispondono alle zone chiare rotonde e risultano ingrandite di 200 000 diametri. Il virus Semliki Forest (SFV), che trae il nome dalla foresta pluviale dell'Uganda dove è stato isolato per la prima volta, infetta molti vertebrati e invertebrati. È affine a virus che provocano l'encefalite, ma non è patogeno per l'uomo. Come altri virus, manca dell'apparato necessario per autoriprodursi. Solo penetrando in una cellula animale ed espropriandone la macchina biosintetica, è in grado di produrre nuove particelle, ciascuna delle quali consiste di una molecola di RNA a filamento singolo, contenuta in un involucro proteico: il capside che è ricoperto da una membrana analoga a quella che protegge la cellula animale. Nella membrana virale sono incluse molecole proteiche, visibili attorno alle particelle come uno strato chiaro di lanugine.**

rivolte verso l'esterno e le code idrofobe verso l'interno, l'una contro l'altra.) Nel contesto della membrana, le molecole lipidiche sono in grado di spostarsi liberamente in senso laterale, ma non verticale. La libertà di movimento in due direzioni dello spazio conferisce alla membrana le proprietà di un liquido e ciò è essenziale per molte funzioni della membrana stessa, tra cui il processo che porta alla liberazione del virus dalla cellula ospite.

Le molecole lipidiche esistono in numerose forme, diverse a seconda dei gruppi chimici che contengono. Queste differenze hanno reso possibile l'identificazione dell'origine delle molecole lipidiche nella membrana del SFV. Ossi Renkonen e Leeri Kääriäinen con i loro collaboratori presso l'Università di Helsinki hanno dimostrato che la composizione lipidica di tale membrana dipende direttamente dalla cellula ospite in cui la particella virale è stata costruita.

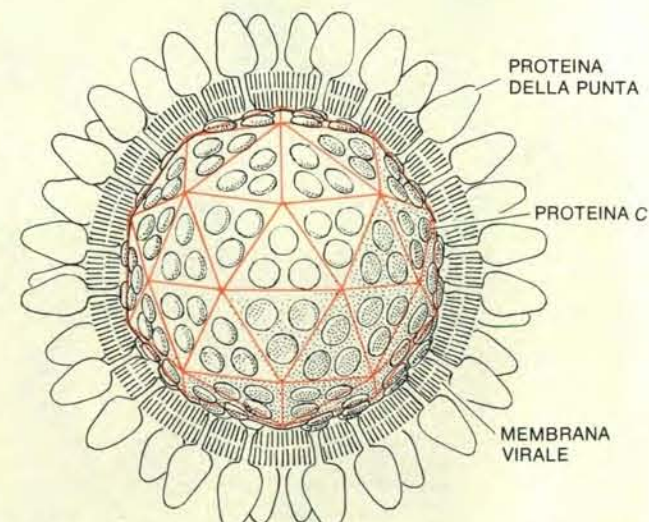
In che modo il virus Semliki Forest penetra in una cellula animale? Nel nostro tentativo di rispondere a questa domanda abbiamo operato soprattutto con cellule renali di criceto, coltivate in laboratorio. I sistemi sperimentali basati sull'uso di cellule in coltura non riescono a spiegare interamente il processo di infezione virale nell'organismo integro; possono, però, aiutare a chiarire come un virus entra in una cellula e ne esce. Benché le nostre conoscenze sugli ultimi stadi dell'infezione virale nelle cellule coltivate fossero abbastanza complete, rimaneva ancora oscuro il modo in cui il materiale genetico del virus si apriva il varco verso il citoplasma della cellula ospite. Doveva trattarsi di un meccanismo attivo, dato che grosse molecole come le proteine e strutture ancora più voluminose come i virus non potevano semplicemente diffondere attraverso la membrana cellulare.

Per spiegare in che modo virus dotati di membrane esterne penetrano nelle cellule animali sono stati proposti due meccanismi. L'ingresso potrebbe avvenire grazie alla fusione della membrana virale con la membrana della cellula e successiva liberazione del nucleocapside nel citoplasma. Una fusione di questo tipo è stata osservata in alcuni virus, ma non è stato chiarito se tale processo può condurre all'infezione. Un meccanismo alternativo potrebbe consistere nella formazione, da parte della membrana cellulare, di una vescicola che avvolga la particella virale, e che successivamente, staccatasi dalla membrana, trasporti la particella nel citoplasma. Per alcuni virus è provato che penetrano nelle cellule grazie a questo processo, chiamato endocitosi; non è stato però dimostrato che l'endocitosi porti all'infezione. La via di ingresso del virus Semliki Forest nella cellula animale, che abbiamo scoperto in collaborazione con Erik Fries, Mark Marsh e Judith White dell'EMBL e Jürgen Kartenbeck del Centro tedesco per le ricerche sul cancro, ha in realtà aspetti che si rifanno ad ambedue i meccanismi.

La nostra ricerca ha anche evidenziato che le proteine delle punte e la membrana del virus non costituiscono un semplice rivestimento protettivo per il nucleocapside, ma svolgono anche un ruolo fondamentale nella liberazione della particella virale nel citoplasma. Il primo stadio dell'ingresso è l'attacco delle glicoproteine delle punte alla superficie cellulare. I nostri risultati indicano che le glicoproteine del SFV si legano alle glicoproteine della membrana cellulare, ivi compresi gli antigeni da trapianto (proteine del sistema immunitario, responsabili del riconoscimento dei tessuti estranei).

In presenza di una cellula animale, le particelle di SFV si legano di preferenza alla superficie dei microvilli, estroflessioni allungate che si protendono dalla superficie cellulare. Dopo essersi legata a un microvillo, la particella virale discende, lungo la parte cilindrica dello stesso, verso il centro della cellula. Nell'arco di pochi minuti, essa viene a trovarsi in una regione specializzata della membrana cellulare, un «pozzetto rivestito», spesso presente alla base dei microvilli. Lo spesso rivestimento si trova sul lato citoplasmico (interno) di tale regione ed è costituito principalmente da clatrina, una proteina di elevato peso molecolare che fu isolata per la prima volta da Barbara M. F. Pearse del Medical Research Council Laboratory of Molecular Biology di Cambridge, in Inghilterra.

Sulla superficie della cellula si formano in continuazione «pozzetti rivestiti» che, successivamente, si ripiegano all'interno, staccandosi e dando origine a «vescicole rivestite». La clatrina che si trovava sul lato interno della membrana ora avvolge la vescicola, costituendo un «guscio» di forma geometrica simile al capside del SFV. Le vescicole rivestite, spostandosi nel citoplasma, perdono la clatrina e si fondono con un endosoma, un grosso vacuolo con superficie esterna liscia. L'endosoma, a sua volta, si fonde con un lisosoma, il vacuolo ricordato prima avente funzione degradativa, che consiste principalmente nel demolire le sostanze della cellula stessa non più necessarie e quelle penetrate in essa per endocitosi, che, per essere utilizzate, devono essere decomposte. L'operazione viene completata da una serie di enzimi digestivi ed è facilitata dal fatto che, nel lisosoma, il $pH$ è inferiore a quello di altre regioni della cellula, e quindi l'ambiente è più acido.

**La struttura della particella del virus Semliki Forest, che ha un diametro di 65 nanometri, è relativamente semplice. L'RNA all'interno del capside (non visibile) è una catena di 12 700 nucleotidi, la cui sequenza fornisce l'informazione genetica necessaria per la sintesi delle proteine virali. Il capside consiste di 180 molecole di una proteina che viene designata come proteina *C*. (L'RNA e il capside insieme formano un nucleocapside.) Il capside ha la forma di un solido regolare con 60 facce e ogni faccia è un triangolo equilatero in cui sono sistemate tre molecole di proteina *C*. La membrana virale esterna ha due strati di molecole lipidiche, con le «teste» idrofobe rivolte all'esterno e le «code» idrofile all'interno. Le molecole lipidiche sono libere di spostarsi lateralmente, dando così alla membrana le proprietà di un liquido. Nella membrana sono inserite 180 «punte», ognuna composta di tre molecole proteiche fra loro legate. Ogni punta attraversa la membrana e si lega a un'unica proteina *C* del capside. Le punte sporgenti danno alla particella del SFV un caratteristico aspetto *(a sinistra)*.**



Per penetrare nella cellula, le particelle del SFV sfruttano in parte un importante ciclo cellulare. Molte delle molecole necessarie per la regolazione o la nutrizione cellulare fanno il loro ingresso nella cellula per endocitosi. È questo il caso delle lipoproteine, degli ormoni polipeptidici, delle vitamine (facenti parte di molecole più grosse) e dei fattori della crescita. Gran parte di queste molecole finisce nel lisosoma, ma quella più attentamente studiata sotto questo aspetto è la lipoproteina serica a bassa densità, il principale vettore del colesterolo nel sangue. Richard G. W. Anderson, Michael S. Brown e Joseph L. Goldstein del Health Science Center dell'Università del Texas a Dallas hanno dimostrato che essa penetra nella cellula per mezzo di un pozzetto rivestito e viene quindi degradata nel lisosoma; successivamente il colesterolo viene utilizzato per la costruzione di nuove membrane cellulari.

La capacità delle cellule animali di inglobare particelle virali per endocitosi è notevole: una cellula renale di criceto ne può incorporare circa tremila al minuto. Si è calcolata una media di 1,3 particelle per ogni vescicola. Pertanto, alla superficie della cellula devono formarsi ogni minuto almeno 2300 vescicole rivestite. Il trasporto di sostanze all'interno mediante questo meccanismo è tanto intenso che la cellula deve riuscire in qualche modo a riportare in superficie i componenti della membrana che formano le vescicole rivestite, altrimenti la membrana si esaurirebbe. È stato dimostrato che l'endocitosi e il ritorno dei componenti della membrana in superficie costituiscono un ciclo permanente, ma per penetrare nella cellula, il virus ne sfrutta solo la prima fase, che si conclude nel lisosoma.

In che modo la particella di SFV sfugge all'azione distruttiva degli enzimi nel lisosoma? Si sapeva già da qualche tempo che i virus animali, penetrati nella cellula per endocitosi, sono poi trasferiti al lisosoma. Si pensava però che il processo conducesse alla distruzione delle particelle virali piuttosto che all'infezione. La prima prova del contrario si è avuta da esperimenti basati su inibitori dei lisosomi. Ci si è serviti di inibitori debolmente basici e solubili nei lipidi, quando sono privi di carica elettrica. Se privi di carica, essi possono attraversare la membrana esterna del lisosoma. Una volta entrati nel lisosoma, però, le loro molecole acquisiscono una carica elettrica a causa dell'ambiente acido del vacuolo. Assunta la carica non possono più riattraversare la membrana del lisosoma e, pertanto, si accumulano al suo interno, diminuendone l'acidità e inibendone la funzione.

Abbiamo scoperto che la somministrazione di questi inibitori lisosomici alle cellule animali interrompe il processo mediante il quale il nucleocapside viene liberato nel citoplasma della cellula ospite. Essi non compromettono l'attacco della particella del SFV alla membrana cellulare né l'ingresso della particella stessa all'interno del lisosoma; impediscono invece al nucleocapside di uscire dal lisosoma e di ritornare nel citoplasma. Ulteriori ricerche effettuate nel nostro laboratorio hanno dimostrato che gli inibitori interferiscono con la fusione della membrana del virus con quella del lisosoma. In assenza di inibitori, infatti, tale fusione permette alle particelle virali di attraversare il lisosoma e di penetrare nel citoplasma.

**Ogni punta della membrana del virus Semliki Forest è composta da tre molecole proteiche attorcigliate l'una all'altra. La molecola designata come *E*3 è una catena di 66 amminoacidi, e si trova completamente all'esterno della membrana virale. Anche la molecola di *E*1 (composta da 438 amminoacidi) e la molecola di *E*2 (composta da 422 amminoacidi) si estendono per la maggior parte all'esterno della membrana virale. Il breve moncone sul versante interno della membrana consta di 31 amminoacidi per la molecola di *E*2 e di due amminoacidi per la molecola di *E*1. Esso àncora la punta a una molecola di proteina *C* del capside. Ognuna delle tre proteine della punta presenta catene laterali di carboidrati: si tratta quindi di glicoproteine. Per modalità di struttura e sintesi, esse sono molto simili a quelle contenute nella membrana della cellula animale.**

La reazione di fusione è chiaramente innescata dall'acidità interna del lisosoma. Quando il $pH$ di un preparato contenente particelle di SFV e frammenti di membrana lipidica libera è stato abbassato fino a un livello simile a quello dell'interno del lisosoma, la membrana del SFV si è immediatamente fusa con le altre membrane, permettendo al nucleocapside di attraversarle. In un'appendice di questo lavoro sono state prodotte, con lipidi simili a quelli della membrana lisosomica, delle vescicole sintetiche in cui è stata successivamente introdotta ribonucleasi, un enzima che demolisce le molecole di RNA. In presenza di particelle virali e abbassando il $pH$ della soluzione si è assistito alla fusione della membrana virale con quella delle vescicole. Microfotografie elettroniche hanno rivelato che le vescicole erano ricoperte di proteine appartenenti alle punte della membrana virale; i nucleocapsidi erano al loro interno e l'RNA virale era sottoposto all'azione demolitrice della ribonucleasi.

Nelle cellule integre, le particelle virali escono dal lisosoma in questo modo. L'acidità interna del lisosoma induce chiaramente una modificazione nella membrana virale, permettendole di fondersi con la membrana lisosomica. Il processo si svolge tanto rapidamente che il nucleocapside viene espulso nel citoplasma della cellula senza venir distrutto dagli enzimi degradativi del lisosoma.

L'esatto meccanismo mediante il quale viene indotta la fusione non è noto. Tuttavia, abbiamo il sospetto che la proteina *E*1 delle punte della membrana virale svolga un ruolo importante. Essa possiede un segmento idrofobo, composto da 16 amminoacidi, nella porzione che sporge dalla membrana virale. Se l'ambiente circostante è acido, tale segmento può scoprirsi e reagire con la membrana lisoso-

mica; le due membrane risultano quindi intimamente giustapposte e la reazione di fusione può così realizzarsi.

Le ricerche effettuate sul SFV hanno individuato, per la prima volta, la via attraverso cui un virus penetra in una cellula animale e il meccanismo di liberazione del nucleocapside nel citoplasma. Il significato del risultato ottenuto apparirà ancora più rilevante se si scoprisse che tale meccanismo è comune a tutti i virus animali. Ricerche che abbiamo recentemente effettuato in collaborazione con Karl Matlin e Hubert Reggio dell'EMBL indicano che i virus influenzali e il virus della stomatite vescicolare (affine al virus che provoca la rabbia) seguono la stessa via. Come il virus Semliki Forest, entrambi sono dotati di una membrana esterna in cui sono inserite proteine. Anche alcune tossine batteriche possono penetrare nel citoplasma attraverso il lisosoma.

In effetti, il lisosoma può risultare, per la cellula animale, un punto di particolare vulnerabilità a molti agenti patogeni. Le indagini sui rapporti tra virus e lisosomi possono pertanto avere importanti implicazioni cliniche. I farmaci normalmente disponibili per la terapia di infezioni virali hanno efficacia limitata e spesso effetti collaterali indesiderabili. Un farmaco che interferisse con la fusione della membrana virale e lisosomica potrebbe benissimo essere un efficace agente antivirale.

La liberazione del nucleocapside da parte del lisosoma completa la prima fase dell'infezione. La seconda fase comprende la duplicazione dell'RNA virale, la sintesi e la modificazione delle proteine virali e l'assemblaggio del nuovo nucleocapside. Questi processi dipendono in larga misura dalla macchina per la sintesi proteica della cellula ospite, rappresentata dai ribosomi e da svariati enzimi e altre molecole. Perché possa aver luogo la prima duplicazione dell'RNA virale, occorre che l'acido nucleico si liberi dal capside. Quando ciò è avvenuto, dal primo segmento dell'RNA virale vengono tradotte parecchie proteine. Una di queste è l'enzima che catalizza la replicazione dell'RNA, chiamato RNA-polimerasi RNA-dipendente, il quale deve essere codificato dal materiale genetico del virus, dato che non esiste nella cellula ospite. Finora, tuttavia, né la liberazione dell'acido nucleico né la sintesi dell'enzima sono stati totalmente compresi.

**In queste microfotografie, scattate nel laboratorio degli autori e in quello di Jürgen Kartenbeck, compaiono il primo e l'ultimo stadio dell'infezione di una cellula animale da parte del SFV. Il primo stadio è l'attacco della particella alla membrana cellulare *(microfotografia in alto)*. Le proteine della punta della membrana virale si attaccano ai recettori per le proteine presenti nella membrana cellulare. I recettori sono molecole proteiche la cui configurazione permette loro di legarsi a un numero limitato di altre proteine. La particella virale tende a legarsi a un microvillo, prolungamento digitiforme della superficie cellulare *(a destra nella microfotografia in alto)*, quindi penetra in un «pozzetto rivestito», così chiamato a causa dello spesso strato proteico che ne riveste la superficie interna. Il pozzetto rivestito si ripiega poi tutt'attorno alla particella e forma una «vescicola rivestita», destinata a trasportare la particella nel citoplasma *(microfotografia al centro)*. Avvenuta la replicazione virale, le nuove particelle abbandonano la cellula ospite, mediante formazione di gemme all'esterno della membrana cellulare *(microfotografia in basso)*.**

**L'ingresso del virus Semliki Forest in una cellula animale esige che la particella virale passi attraverso parecchi vacuoli. La vescicola è rivestita da una proteina chiamata clatrina, che forma, attorno alla particella, un guscio analogo al capside che racchiude l'RNA virale. All'interno della cellula, la vescicola perde la clatrina e si fonde con un grosso vacuolo dalla superficie liscia: l'endosoma. Molte molecole coinvolte nel normale metabolismo cellulare penetrano nella cellula grazie a questo processo, chiamato endocitosi. L'endosoma poi, a sua volta, si fonde con il vacuolo detto lisosoma. Il lisosoma contiene enzimi che demoliscono le sostanze assunte dalla cellula; molte di queste vengono pertanto trasformate in una forma utilizzabile. L'interno del lisosoma ha un grado di acidità superiore rispetto alle altre parti della cellula. È noto da tempo che le particelle virali possono penetrare nella cellula per questa via; tuttavia, si presumeva che esse venissero demolite dagli enzimi del lisosoma e che non potessero raggiungere il citoplasma per dar inizio all'infezione. Oggi si sa che l'acidità del lisosoma provoca invece una fusione della membrana virale esterna con la membrana del lisosoma stesso. Come conseguenza, il nucleocapside viene liberato nel citoplasma. È la prima volta che si è riusciti a seguire la via percorsa da un virus in una cellula animale fino al punto in cui ha inizio l'infezione. La scoperta può assumere notevole importanza se si riesce a dimostrare che altri virus animali provocano l'infezione allo stesso modo.**

Quando nel citoplasma sono presenti sia l'acido nucleico liberato sia molecole di polimerasi, ha inizio la replicazione. Vengono prodotti, in notevoli quantitativi, due differenti tipi di molecola di RNA. Il primo è una copia della molecola completa dell'acido nucleico virale, composta da 12 700 nucleotidi, e viene designato come RNA 42*S*. (*S* è la sigla per l'unità Svedberg, misura della massa della molecola.) Le molecole 42*S* che si accumulano servono come base per la produzione di una maggior quantità di RNA-polimerasi, e alla fine vengono incorporate in nuovi nucleocapsidi.

Il secondo tipo di RNA è una copia dell'ultimo terzo dell'intera molecola di RNA virale, a partire dalla fine del segmento che codifica per l'RNA-polimerasi. La molecola è più corta, essendo formata da 4170 nucleotidi, ed è indicata come RNA 26*S*. Essa incorpora l'informazione genetica necessaria per la sintesi delle quattro proteine del virus. Le sequenze che codificano per queste proteine sono disposte, lungo il filamento di RNA 26*S*, nel seguente ordine: *C*, *E*3, *E*2 ed *E*1. Nell'arco di quattro ore dal momento in cui il virus viene messo in presenza di una cellula animale, le due forme di RNA virale costituiscono la maggior parte delle molecole di RNA messaggero presenti nella cellula, che si è praticamente trasformata in una catena di montaggio per la fabbricazione di nuove particelle virali.

La sintesi delle proteine del capside virale e delle punte ha inizio quando l'RNA 26*S* si attacca a un ribosoma. Le quattro proteine strutturali vengono tradotte una dopo l'altra a partire da esso. La traduzione comincia in corrispondenza del codone «d'inizio», una tripletta di nucleotidi che precede la sequenza che codifica per la proteina *C*. Il processo continua ininterrotto fino a raggiungere il codone «d'arresto», alla fine del segmento che codifica per la proteina *E*1. La porzione di acido nucleico virale compresa fra il codone «d'inizio» e il codone «d'arresto» si compone di 3760 nucleotidi.

Nella cellula vivente, le quattro proteine virali non vengono solo liberate dal ribosoma nel citoplasma sotto forma di catene libere. La nuova proteina *C*, la prima a essere tradotta, si trova nel citoplasma; la proteina *E*1 e una proteina intermedia nota come *p*62, composta dalle proteine *E*2 ed *E*3 legate fra loro, si trovano invece inserite nella membrana del reticolo endoplasmatico, una rete di membrane interconnesse che svolge una funzione cardine nella sintesi e modificazione delle proteine della cellula ospite, comprese le proteine della membrana e le proteine che la cellula secerne. La produzione delle proteine della membrana virale è quasi identica a quella delle proteine della membrana della cellula ospite.

La struttura delle glicoproteine nella membrana delle cellule animali e il modo in cui tali molecole sono sintetizzate sono abbastanza noti. Molte glicoproteine cellulari presentano un caratteristico gruppo di 20-30 amminoacidi all'estremità della catena proteica, nota come estremità N-terminale, che è la prima parte della catena a essere sintetizzata sul ribosoma. Questo segmento, detto «sequenza segnale», ha una particolare affinità per un sito presente sulla superficie della membrana del reticolo endoplasmatico. Non appena i 20-30 amminoacidi sono stati tradotti da un ribosoma che si sposta liberamente nel citoplasma, il ribosoma stesso si attacca al reticolo endoplasmatico e quindi procede con la traduzione dell'RNA messaggero, cioè di quell'RNA che trasporta l'informazione genetica dal DNA ai ribosomi. Tuttavia, la catena di amminoacidi non si allunga nel citoplasma: si fa strada attraverso la membrana del reticolo endoplasmatico e si estende poi all'interno del reticolo stesso.

Con un esperimento messo a punto da Günter Blobel e da Bernhard Dobberstein della Rockefeller University, abbiamo potuto dimostrare che lo stesso processo interviene nella sintesi delle proteine della membrana del virus Semliki Forest. Abbiamo operato con un sistema privo di cellule e comprendente ribosomi, enzimi, RNA di trasferimento, amminoacidi e altre sostanze necessarie per la sintesi proteica. In assenza del reticolo endoplasmatico, il sistema non sintetizzava correttamente le proteine delle punte della membrana virale. Aggiungendo però le vescicole del reticolo endoplasmatico, tali proteine venivano sintetizzate proprio come all'interno di una cellula infettata dal virus. E si trattava, per la precisione, delle tre proteine *C*, *E*1 e *p*62. Per di più, le catene di amminoacidi si collocavano al posto giusto: la proteina *C* all'esterno delle vescicole nella regione corrispondente al citoplasma della cellula intatta, mentre le proteine *E*1 e *p*62 erano inserite nella membrana della vescicola.

Come è risultato da questi e da altri esperimenti, subito dopo essere stata tradotta, nella cellula intatta, la proteina *C* viene staccata dalla catena, che continua ad allungarsi, a opera di una proteasi, un enzima che taglia appunto le proteine. Tale distacco presenta una sequenza segnale all'estremità della catena della *p*62: essa lega il ribosoma al reticolo endoplasmatico e innesca così l'introduzione della catena proteica nello spessore della membrana. Quasi tutta la catena *p*62 viene allora tradotta a partire dall'acido nucleico virale e migra poi all'interno del reticolo endoplasmatico. Tale spostamento viene

chiaramente bloccato dal gruppo di amminoacidi idrofobi, cioè vicini all'estremità della molecola $E2$. Non potendo passare attraverso la membrana lipidica parimenti idrofoba, ancorano la catena proteica al reticolo endoplasmatico. Quando la traduzione della $p62$ si è conclusa, la catena si fissa alla membrana, e 31 amminoacidi si protendono nel citoplasma mentre il resto della catena si trova all'interno del reticolo endoplasmatico.

La seconda proteina che viene sintetizzata è la $E1$. Naturalmente ha una propria sequenza segnale, codificata da un segmento dell'RNA virale compreso tra i tratti che corrispondono alle proteine $E1$ ed $E2$. Questa sequenza, designata come $6K$, è composta da 60 amminoacidi. Viene rimossa dalla catena $E1$ in allungamento mediante due tagli operati da enzimi catalizzatori, dopo che la catena stessa si è legata alla membrana del reticolo endoplasmatico. Il passaggio della proteina $E1$ attraverso la membrana, così come quello della proteina $p62$, viene bloccato da una sequenza di amminoacidi idrofobi prossima all'estremità della catena. Questa rimane ancorata al reticolo endoplasmatico, mentre un breve segmento si protende nel citoplasma. La traduzione della proteina $E1$ è però meno nota rispetto agli altri processi che intervengono nella sintesi delle proteine virali.

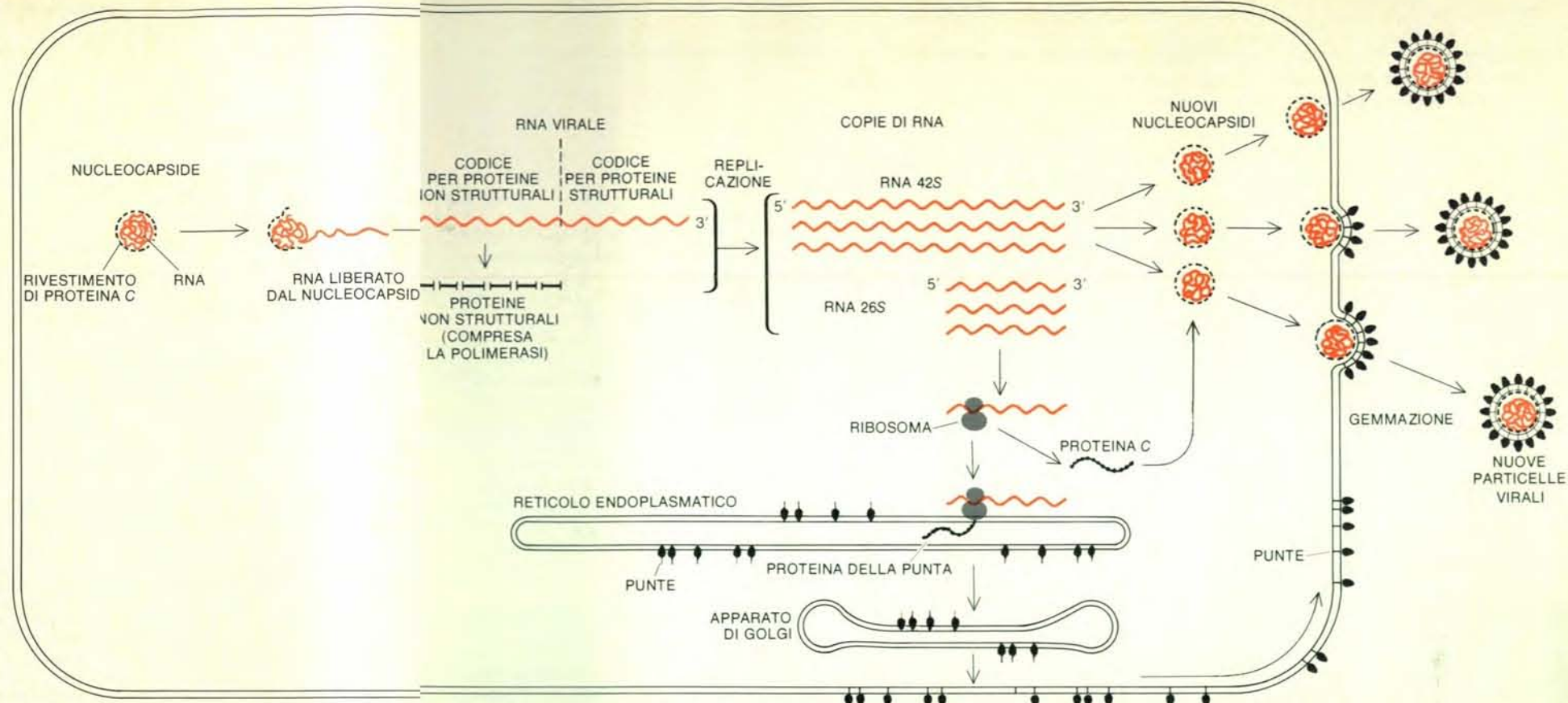
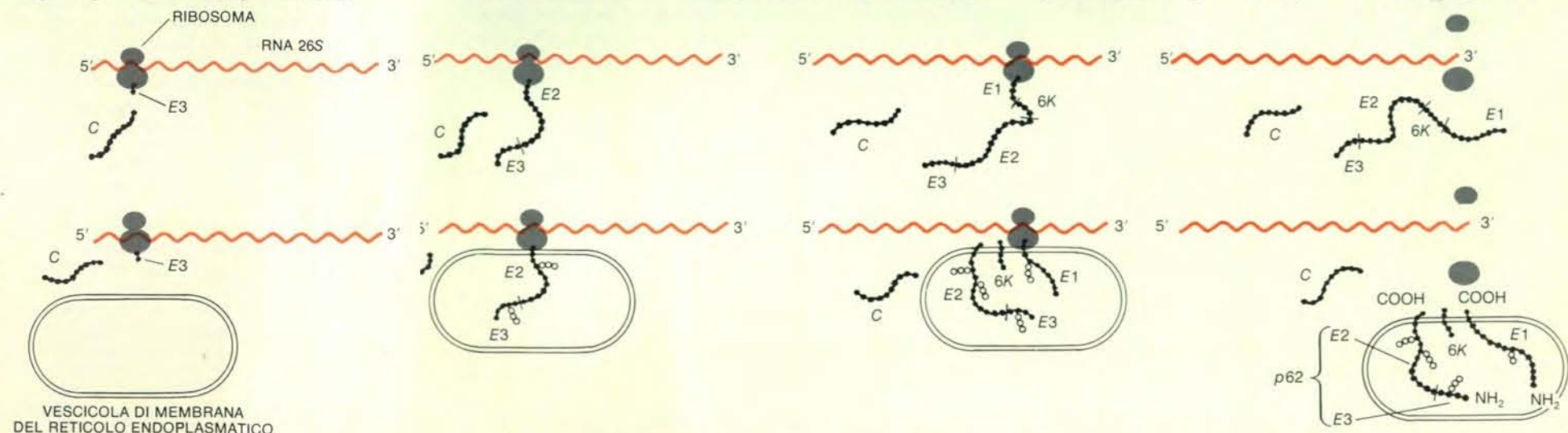
Le catene laterali di carboidrati delle glicoproteine virali vengono attaccate nel corso della traduzione e, più tardi, subiscono alcune modificazioni chimiche. Lo stesso tipo di processo si svolge nella sintesi di alcune glicoproteine proprie della cellula. E, in effetti, la traduzione delle proteine virali non dipende da meccanismi esclusivi. L'unico carattere insolito del processo è la localizzazione delle sequenze segnale che innescano l'attacco del ribosoma al reticolo endoplasmatico. Per la maggior parte delle glicoproteine di cellule animali, la sequenza segnale è codificata da nucleotidi posti immediatamente dopo il codone d'inizio; le sequenze segnale delle proteine virali $E1$ e $p62$ vengono invece codificate da nucleotidi che si trovano all'interno della catena di RNA messaggero. Tuttavia, troppo poche sono le molecole di RNA messaggero caratterizzate completamente perché si possa escludere la possibilità che alcune di esse contengano regioni interne che codifichino sequenze segnale.

La dipendenza del virus dai normali processi cellulari continua nella fase in cui le proteine delle punte della membrana virale vengono modificate, trasportate e inserite nella membrana cellulare. Le proteine $E1$ e $p62$, subito dopo la sintesi, vengono legate fra loro nella membrana del reticolo endoplasmatico. Il complesso che ne risulta viene trasportato alla superficie della cellula per la stessa via seguita dalle proteine di secrezione e da alcune proteine della membrana: dal reticolo endoplasmatico le proteine delle punte si spostano all'apparato di Golgi, un organello complesso che presenta parecchie lacune interne delimitate da membrane, e qui le catene laterali di carboidrati vengono modificate. Michael F. G. Schmidt e Milton J. Schlesinger della Washington University School of Medicine hanno dimostrato che, nell'apparato di Golgi, a ogni proteina vengono legate molecole dotate di lunghe catene laterali idrofobe. È noto che tali molecole sono attaccate anche alle glicoproteine della membrana cellulare e alle proteine della membrana di altri virus, ma il significato di questa modificazione permane oscuro.

Nel suo passaggio attraverso l'apparato di Golgi, il complesso formato dalle due proteine $E1$ e $p62$ conserva un unico orientamento, che è opposto a quello che avrà nella membrana virale. Il segmento destinato a formare la punta sporgente della membrana virale si trova all'interno dell'apparato di Golgi; il segmento che attraverserà la membrana virale è inserito nelle membrane dell'apparato; infine i brevi segmenti delle molecole proteiche che si attaccheranno al capside fuoriescono dall'apparato, penetrando nel citoplasma.

Dopo essere passate attraverso l'apparato di Golgi, le glicoproteine collegate vengono portate alla superficie della cellula ospite e quindi inserite nella membrana cellulare, ma ben poco si sa del modo in cui si svolge tale spostamento prima all'interno e poi all'esterno dei suddetti organelli. La scissione della proteina $p62$ nelle catene $E2$ ed $E3$ ha luogo all'incirca quando il complesso raggiunge la superficie cellulare, e avviene a opera di una proteasi. Molte proteine di secrezione, tra cui gli ormoni e le proteine ematiche, vengono trasformate nella loro forma definitiva in un modo analogo.

Allorché il complesso delle glicoproteine delle punte è stato inserito nella membrana cellulare, nel citoplasma - a partire dalla proteina $C$ e dall'acido nucleico - avviene l'assemblaggio del nucleocapside. La fase finale del passaggio del SFV attraverso la cellula animale può quindi cominciare: si forma la particella virale completa che viene poi liberata dalla cellula. Nicholas H. Acheson e Igor Tamm della Rockefeller University hanno dimostrato per la prima volta che la particella di SFV viene montata e poi eliminata dalla cellula mediante gemmazione della membrana cellulare. Una porzione di membrana si avvolge attorno al nucleocapside e, infine, la gemma si stacca dalla cellula. La maggior parte degli altri virus dotati di membrane, tra cui i virus influenzali, il virus del morbillo e quello della rabbia, vengono anch'essi liberati per gemmazione dalla cellula ospite.

La chiave per comprendere l'assemblaggio finale del virus Semliki Forest è stata trovata quando abbiamo osservato che le proteine virali attraversano la membrana e si protendono nel citoplasma. Nella membrana cellulare, le proteine delle punte assumono lo stesso orientamento che assumeranno nella membrana virale, con la punta sporgente all'esterno e i brevi monconi delle proteine $E1$ ed $E2$ che si protendono all'interno della

La produzione dei componenti della particella del virus Semliki Forest si basa sull'apparato di sintesi proteica della cellula. La replicazione ha inizio subito dopo la liberazione del nucleocapside da parte del lisosoma. L'RNA virale perde il suo involucro proteico; immediatamente dopo, viene prodotta una RNA-polimerasi specifica per quel virus. (L'RNA-polimerasi è l'enzima che catalizza la replicazione dell'RNA.) La polimerasi e l'RNA si combinano, dando luogo alla produzione di due tipi di nuove molecole di RNA virale: l'RNA 42*S*, che è una copia della molecola completa di RNA virale e l'RNA 26*S*, una molecola più corta, che incorpora il codice genetico per le quattro proteine virali. (*S* sta per Svedberg, unità di misura della massa della molecola.) Nell'arco di tre o quattro ore, questi due tipi di molecole rappresentano le forme più abbondanti di RNA nella cellula. A questo punto, l'RNA 26*S* si attacca a un ribosoma e comincia ad agire da RNA messaggero per le proteine virali. Le quattro proteine vengono prodotte senza sosta seguendo l'ordine $C$, $E3$, $E2$ ed $E1$. Appena prodotta la proteina $C$ viene staccata ad opera di un enzima e, nel citoplasma, si congiunge con una molecola di RNA 42*S*, formando un nuovo nucleocapside. Successivamente viene prodotta la proteina $E3$. Subito dopo l'unione dei primi pochi amminoacidi della $E3$, il ribosoma si attacca al reticolo endoplasmatico, costituito da una rete di membrane interconnesse all'interno della cellula. La catena di $E3$ in fase di allungamento attraversa la membrana del reticolo endoplasmatico e l'assemblaggio prosegue, con il prolungamento della catena, all'interno del reticolo. Una volta sintetizzate, le tre molecole proteiche della punta vengono combinate per formare la punta della membrana e quindi trasportate all'organello conosciuto come apparato di Golgi, dove vengono modificate. La punta si inserisce nella membrana cellulare, dove incontra un nuovo nucleocapside. Alla fine, la particella virale completa viene espulsa dalla cellula per gemmazione.

Se si pongono molecole di RNA 26*S* in una soluzione contenente ribosomi, enzimi e le altre sostanze necessarie per la sintesi proteica, si ha la formazione di due catene di amminoacidi: la molecola della proteina $C$ e una lunga catena costituita dalle molecole $E3$, $E2$ e $E1$. Questo non corrisponde però al modo in cui le proteine della punta vengono sintetizzate nella cellula ospite. Ma quando alla soluzione vengono aggiunte vescicole fatte di membrane del reticolo endoplasmatico, le proteine virali vengono sintetizzate come se fossero nella cellula. La proteina $C$ viene prodotta e poi staccata dalla catena. I primi amminoacidi della catena $E3$ costituiscono una «sequenza segnale», che lega il ribosoma alla vescicola. La traduzione procede con la catena $E3$, che attraversa la membrana. Le molecole $E3$ ed $E2$ vengono prodotte una dopo l'altra e rimangono unite a formare una proteina nota come $p62$. La traduzione procede ancora fino all'estremità della catena $p62$. A questo punto un segmento di amminoacidi idrofobi impedisce alla catena di attraversare la membrana; essa rimane fissata a quest'ultima mentre la traduzione prosegue. Quando la traduzione della $p62$ è completa, la catena proteica si stacca e comincia la sintesi della molecola $E1$, con una propria sequenza segnale, $6K$, che in seguito viene staccata. L'assemblaggio della molecola $E1$ procede fino a che il passaggio attraverso la membrana non è bloccato da un segmento idrofobo. La proteina $C$ si pone allora fuori della vescicola, nella regione corrispondente al citoplasma. Il tratto maggiore delle catene $p62$ ed $E1$ rimane all'interno della vescicola, nella regione che corrisponde all'interno del reticolo endoplasmatico. Se le proteine vengono prodotte all'interno di una cellula, la catena $p62$ viene tagliata per formare le catene $E2$ ed $E3$; le tre proteine vengono poi assemblate a formare la punta.

**La molecola di RNA 26*S* porta l'informazione genetica necessaria alla sintesi delle quattro proteine virali, indicando al ribosoma quali amminoacidi siano da aggiungere alla catena in allungamento. I siti di glicosilazione rappresentano i punti della proteina in cui si attaccheranno le catene laterali di carboidrati. La sequenza segnale riconosce il punto di attacco del ribosoma del reticolo endoplasmatico. Molte molecole di RNA messaggero nelle cellule animali possiedono regioni che codificano per sequenze segnale. Negli RNA esaminati in dettaglio esse si trovano alla fine della molecola dell'RNA, piuttosto che al centro, come sono invece le sequenze *E*3 e 6*K*. Pochi sono gli RNA messaggeri delle cellule animali studiati tanto accuratamente da poter determinare il loro intero codice genetico; alcuni potrebbero contenere regioni interne codificanti per le sequenze segnale, come accade per l'RNA virale. A parte questa possibile anomalia, la sintesi delle proteine virali segue fedelmente lo schema di molte proteine della cellula stessa.**

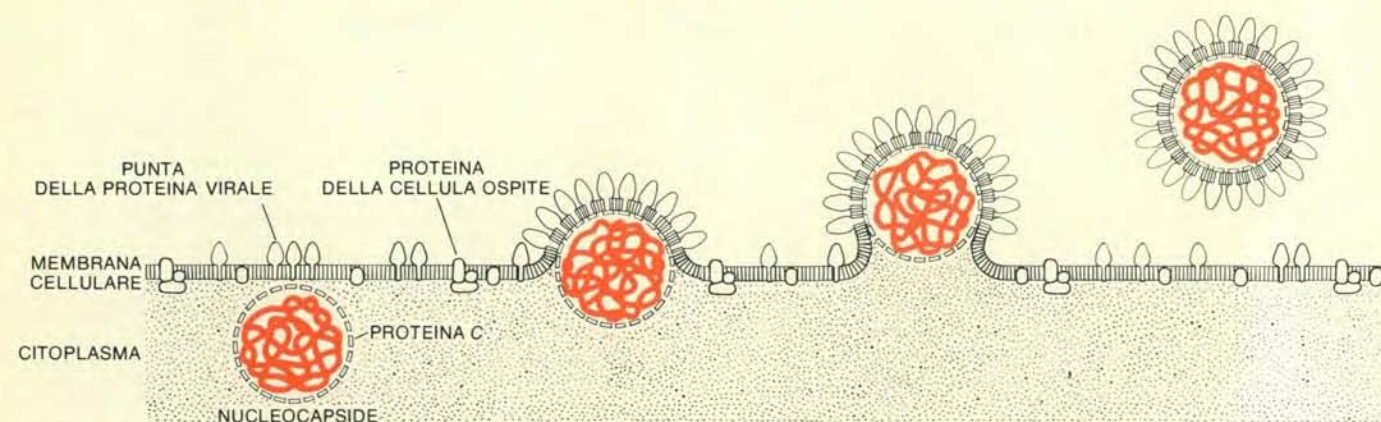
cellula. Le nostre ricerche ci portano a ritenere che la forza motrice del processo di gemmazione derivi dall'interazione chimica dei monconi con la proteina *C*.

La gemmazione comincia quando il nucleocapside si diffonde dal punto dove, nel citoplasma, viene prodotto a un punto situato immediatamente sotto la membrana cellulare. Secondo la nostra ipotesi, una molecola di proteina *C* si lega ai monconi di un'unica punta e tale legame ancora il nucleocapside alla membrana. Il nucleocapside, a sua volta, serve come punta di attacco per altre punte. Dato il carattere di fluido bidimensionale della membrana cellulare, sia le glicoproteine virali sia quelle cellulari sono libere di spostarsi in senso laterale. Quando una punta si sposta sopra l'area dove è ancorato il nucleocapside, si lega a una proteina *C* vacante. Formandosi nuovi legami tra proteine delle punte e proteine *C*, la membrana cellulare è costretta a incurvarsi attorno al nucleocapside e, quando tutte le 180 molecole del capside si sono attaccate alle punte, finisce per avvolgerlo del tutto. La membrana poi si rompe e la particella virale viene a trovarsi libera nello spazio intercellulare. Dopo il distacco di tale particella, nella membrana non rimane però alcuna lacerazione in quanto, come tutte le membrane lipidiche, è in grado di risaldarsi spontaneamente.

Il meccanismo da noi proposto mostra chiaramente che i lipidi che compongono la membrana virale sono uguali a quelli della membrana della cellula ospite. Non è altrettanto chiaro, invece, perché nessuna glicoproteina della membrana della cellula ospite sia presente nella membrana virale. Può darsi che le proteine dell'ospite siano allontanate dal punto in cui avviene la gemmazione perché la loro affinità chimica per le proteine *C* è inferiore a quella delle glicoproteine virali.

Poche ore dopo l'ingresso del virus Semliki Forest in una cellula animale, sul lato esterno della membrana cellulare cominciano a formarsi per gemmazione nuove particelle virali. Una volta che queste si trovano nello spazio intercellulare, si allontanano dirigendosi verso altre cellule e diffondono in tal modo l'infezione. Se si considera il ciclo virale dal momento dell'ingresso a quello della liberazione, risulta sorprendente la scarsità di meccanismi propri del virus. Il legame della particella virale alla superficie cellulare e la fusione della sua membrana con la membrana del lisosoma dipendono dalle proteine della membrana virale; una RNA-polimerasi specifica per il virus catalizza la replicazione dell'acido nucleico virale; l'uscita dalla cellula richiede l'interazione delle proteine della punta della membrana virale con la proteina *C* virale.

Per il resto, il virus si basa sui normali processi della cellula animale. Penetra nella cellula per una via che è seguita da molte macromolecole necessarie al metabolismo. Il passaggio dal ribosoma al reticolo endoplasmatico, all'apparato di Golgi, alla membrana cellulare corrisponde a quello di numerose proteine cellulari. La sintesi delle molecole di RNA virale e delle quattro proteine strutturali virali sfrutta l'apparato di sintesi proteica che opera normalmente nella cellula animale.

La pesante dipendenza dei virus, e quindi anche del virus Semliki Forest, dalle normali funzioni della cellula ospite permette di comprendere come mai gli agenti antivirali siano meno efficaci di quelli antibatterici. Non è semplice, infatti, preparare un farmaco che interferisca con la replicazione del virus senza danneggiare simultaneamente la cellula ospite. Ricerche tese alla produzione di utili agenti antivirali dovrebbero probabilmente cercare di bloccare le poche funzioni specifiche del virus. D'altra parte, la pesante dipendenza del virus dalle normali funzioni cellulari implica che le ricerche sui virus forniscano indicazioni non solo per la pratica clinica, ma anche per una più profonda comprensione della vita della cellula animale.

**La liberazione della particella di virus Semliki Forest dalla cellula ospite avviene in seguito alla formazione di gemme all'esterno della membrana cellulare. Il processo ha inizio con lo spostamento di un nuovo nucleocapside in una zona immediatamente sotto la membrana. Una proteina *C* si lega al corto moncone della punta della membrana, ancorando così il nucleocapside alla membrana. Legami analoghi si formano tra altre molecole di proteina *C* e altre punte. Le punte possono liberamente spostarsi in senso laterale nella membrana e sembrano avere una maggiore affinità chimica per le proteine *C* piuttosto che per le proteine della membrana cellulare. Pertanto queste ultime vengono allontanate dal nucleocapside. Con il procedere del legame, la membrana cellulare si incurva attorno ad esso. Quando tutte le 180 molecole di proteina *C* sono legate alle punte, la membrana risulta tutta avvolta attorno al nucleocapside e si rompe. La particella virale si libera così e penetra nello spazio intercellulare per rinnovare il processo di infezione. Il meccanismo di uscita spiega come mai le molecole lipidiche della membrana virale siano identiche a quelle della cellula ospite mentre le proteine sono specifiche del virus Semliki Forest.**

# Le orme fossili di Laetoli

*In questo sito, in Tanzania, conservate da ceneri vulcaniche depositatesi tre milioni e mezzo di anni orsono, sono state rinvenute migliaia di orme animali e qualche orma di ominide*

di Richard L. Hay e Mary D. Leakey

Nei pressi del lago Eyasi, in Tanzania, esiste una serie di strati di ceneri vulcaniche grazie ai quali sono stati preservati resti di ominidi fra i più antichi che si conoscano, in quanto risalgono a ben 3,5-3,8 milioni di anni fa. Questi strati di ceneri hanno conservato inoltre qualcosa di ancor più insolito: orme fossili. In questi depositi sono state infatti scoperte varie decine di migliaia di orme di animali. La sopravvivenza di queste tracce, normalmente effimere, che vanno dalle impronte lasciate dai piedi di antichi ominidi alla traccia del passaggio di un insetto, ci consente di gettare uno sguardo di insolita immediatezza sulla vita della savana africana molto prima del Pleistocene. Ma in che modo queste impronte hanno potuto sottrarsi alla cancellazione?

L'estesa formazione nota ai geologi con il nome di strati di Laetoli copre una superficie di circa 1500 chilometri quadrati sull'altopiano di Eyasi, un blocco di faglia sollevato a nord-ovest del lago Eyasi. I depositi poggiano su antiche rocce di basamento del Precambriano e sono attorniati e sovrastati verso est da vari grandi vulcani. Le zone esposte più ricche di fossili sono situate nell'area più ristretta chiamata Laetoli, di circa 70 chilometri quadrati a un'altitudine di 1700-1800 metri in prossimità dello spartiacque che corre fra il lago Eyasi e, più a nord, la Gola di Olduvai.

I fossili sono distribuiti soprattutto nei 45-60 metri superiori dei depositi, che a Laetoli raggiungono uno spessore di almeno 130 metri. Circa tre quarti della parte superiore della formazione sono costituiti da tufi eolici: strati di ceneri vulcaniche che furono ridepositate dal vento dopo la loro caduta. La maggior parte degli altri strati di cenere, che si alternano ai tufi eolici, sono sedimenti di cenere rimasti essenzialmente indisturbati dopo la ricaduta dalla nube eruttiva. Le ceneri eoliche, formate in gran parte da particelle delle dimensioni di minuscoli granelli di sabbia (da 0,125 a 0,25 millimetri) e di granelli di sabbia di medie dimensioni (da 0,25 a 0,5 millimetri), ricoprirono ossa e denti di animali, uova di uccelli, chiocciole di terra e altri oggetti esposti sul suolo. Tutte le ceneri, indisturbate ed eoliche, provenivano da un vulcano, il Sadiman, a circa 20 chilometri a est di Laetoli.

Nel 1935 quest'area fu esplorata da Louis Leakey, dal geologo Peter Kent e da uno di noi (Mary Leakey). Il gruppo trovò che i resti animali fossili a Laetoli erano più antichi di quelli scoperti nel livello più basso (strato I) della Gola di Olduvai. Nel 1938-1939 una spedizione, guidata da Ludwig Kohl-Larsen, dell'Università di Tübingen, raccolse in quest'area un gran numero di fossili, tra cui un frammento di mascella superiore di un ominide. Nel 1974 uno di noi (Mary Leakey) rinvenne altri resti di ominidi in un migliore stato di conservazione.

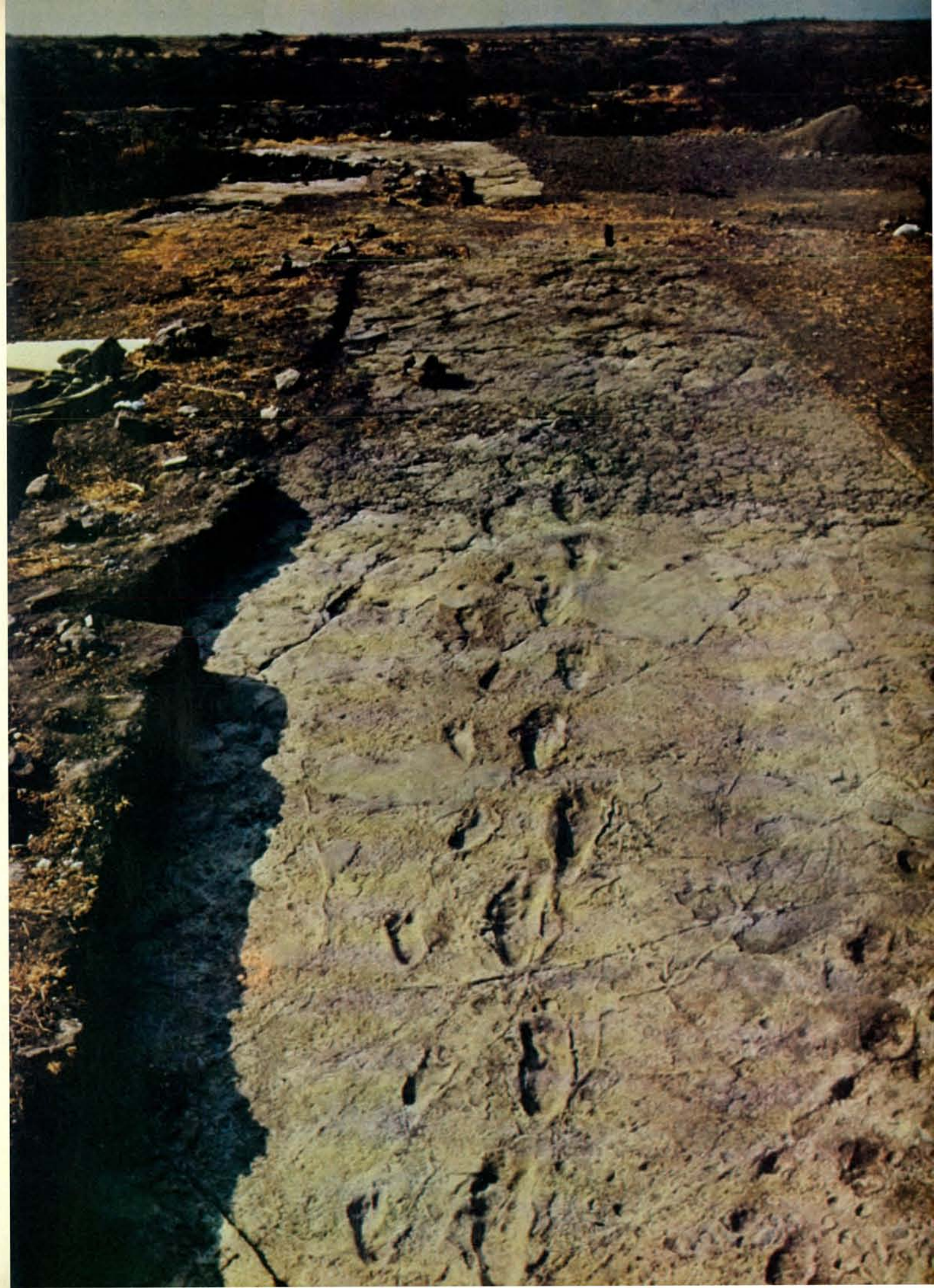
Indagini sistematiche dell'area, condotte dal 1975 sino a oggi, hanno portato in luce numerosi resti fossili di ominidi, fra cui principalmente mandibole e denti, e una ricca varietà di altri resti animali.

Nel 1976 Andrew Hill, della Harvard University, si imbattè per la prima volta in orme di animali in uno strato di tufo che, da allora, è stato chiamato il Tufo delle Orme (Footprint Tuff). L'anno seguente Peter Jones dell'Università di Oxford e Philip Leakey trovarono, nello stesso strato, ma in peggiori condizioni di conservazione, orme lasciate probabilmente da ominidi e nel 1978 Paul I. Abell, dell'Università di Rhode Island, scoprì nel tufo, in un altro sito, un'orma inequivocabilmente di ominide. La rimozione dello strato superficiale rivelò altre orme di ominidi in due lunghe tracce parallele.

Molti di questi strati di tufo contengono il minerale biotite, ricco di potassio. Ciò rende possibile la loro datazione col metodo del potassio-argo, che si fonda sul decadimento dell'isotopo radioattivo potassio 40 nel gas nobile argo. Lavorando su campioni di biotite prelevati da uno strato sottostante ai depositi fossiliferi più profondi del sito e su altri campioni della parte superiore di questa formazione, Garniss H. Curtis e Robert Drake, dell'Università della California a Berkeley, delimitarono col metodo potassio-argo l'età del Tufo delle Orme. I campioni prelevati dagli strati inferiori indicarono che il tufo aveva un'età di 3,8 milioni di anni; l'età dei campioni tratti dal Tufo delle Orme e dalla zona sovrastante copriva un intervallo di tempo molto ristretto attorno a 3,5 milioni di anni.

In generale gli animali conservati allo stato fossile a Laetoli sono simili per tipo a quelli viventi oggi nella stessa area. In ordine di quantità decrescente, i resti di vertebrati più comuni sono quelli di bovidi (antilopi e forme affini), lagomorfi (e precisamente conigli), giraffe, rinoceronti, cavalli, maiali e due tipi di proboscidati (elefanti e dinoteri, una forma oggi estinta). Significativamente assenti sono i resti di coccodrilli, di ippopotami e di altri animali di abitudini acquatiche.

I tufi indisturbati che si alternano con i tufi eolici negli strati di Laetoli ammontano a una cinquantina. Molti di essi attraversano l'intera area ricca di fossili e possono quindi essere utilizzati per stabilire correlazioni stratigrafiche. La maggior parte dei tufi indisturbati hanno uno spessore compreso fra uno e 10 centimetri e non presentano laminazione interna: ciascun tufo fu evidentemente il prodotto di una singola eruzione del Sadiman. Tre di tali tufi presentano invece una laminazione. Il loro spessore varia fra i 12 e i 30 centimetri e rappresentano evidentemente il risultato di una serie di eruzioni ravvicinate. Mentre le

**Queste tracce parallele di impronte di ominidi si estendono per 25 metri attraverso una superficie di cenere vulcanica cementata nel Sito *G*, una delle 16 località di Laetoli, dove numerose tracce di vari animali sono impresse nell'antica serie di strati di cenere comunemente noti come Tufo delle Orme. Le tracce a destra di quelle di ominidi sono quelle di un cavallo a tre dita estinto, lo *Hipparion*. La sequenza di orme di ominide a sinistra fu prodotta dal più piccolo dei tre individui dotati di locomozione eretta; l'altra fu lasciata da un individuo più grande, ma le sue orme furono in parte cancellate da un terzo ominide più piccolo, che ha posato i piedi sulle sue impronte. Non possiamo dire se le serie di orme sono state lasciate contemporaneamente o in tempi diversi. La fotografia è riprodotta per gentile concessione di John Reader, © National Geographic Society.**

Questo ombrellone portatile offrì un po' di sollievo a Mary Leakey *(a sinistra)*, a R. J. Clarke, del National Museum di Bloemfontein *(al centro)* e a due tecnici, mentre nel 1979, stavano liberando le tracce degli ominidi, scoperte nel 1978 nel Sito *G*, dal tufo sovrastante.

Queste righe parallele, lunghe circa sei centimetri, attestano il passaggio di un insetto, forse uno scarabeo stercorario, sulla cenere di Laetoli.

Orma di zampa, lunga 3,5 centimetri, lasciata da un piccolo felino. Si tratta di un'altra fra le migliaia di impronte scoperte a Laetoli.

particelle di cenere che formano i tufi eolici hanno dimensioni comprese fra quelle della sabbia fine e della sabbia media, le particelle dei tufi formati da ceneri indisturbate hanno dimensioni comprese fra quelle della sabbia fine e quelle della sabbia grossa (da 0,5 a 1 millimetro). Le particelle, formate da frammenti di lava vetrosa, hanno un basso tenore di silice e contengono un minerale ricco di calcio, la melilite.

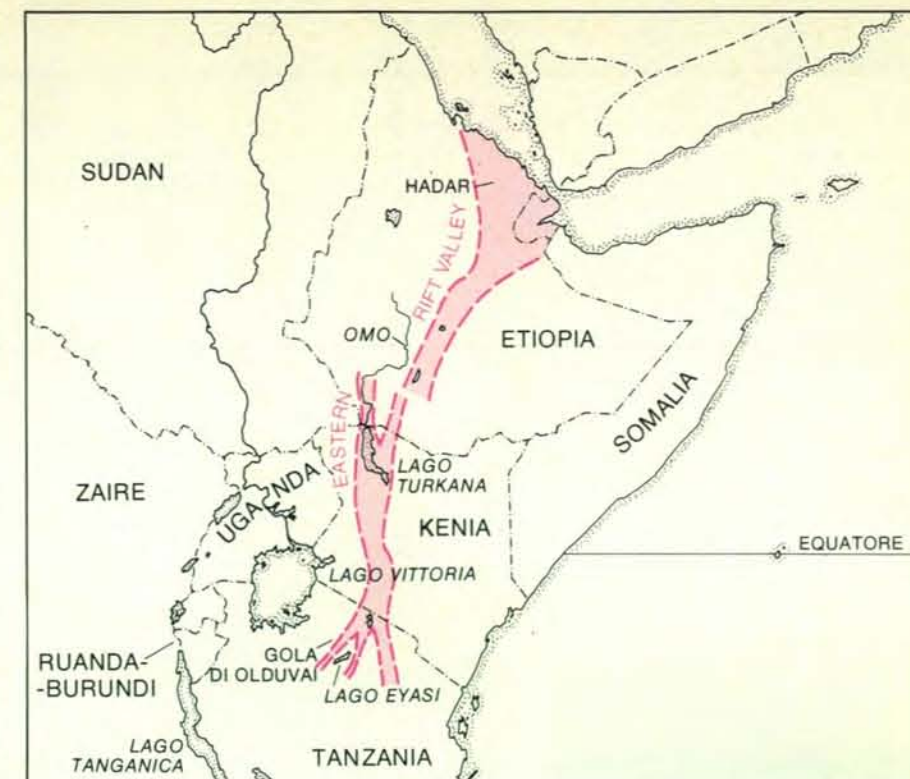
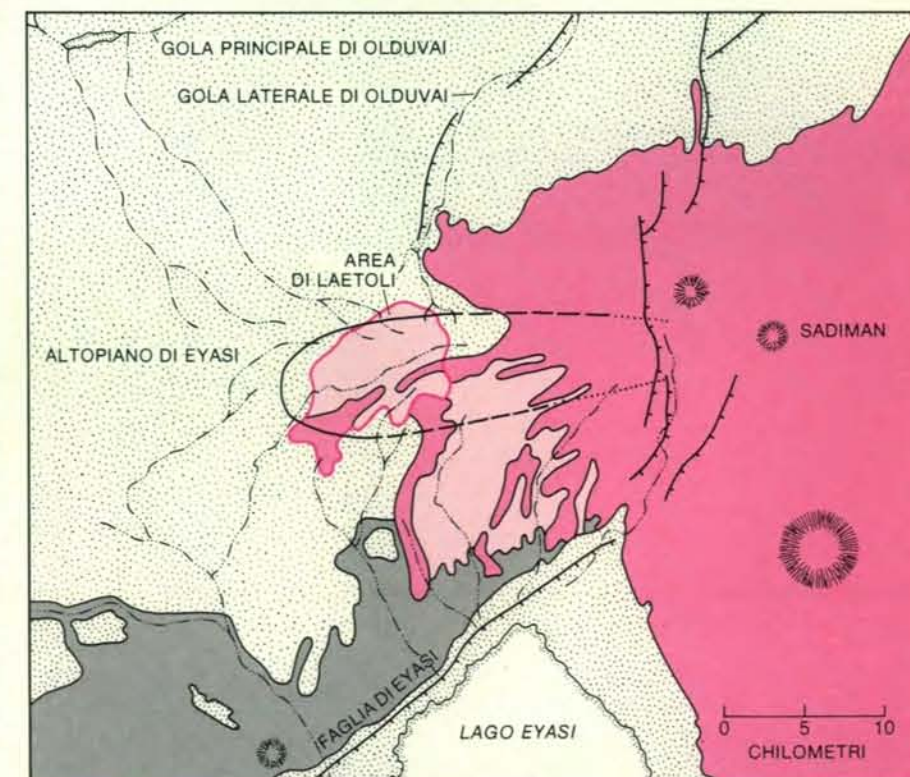
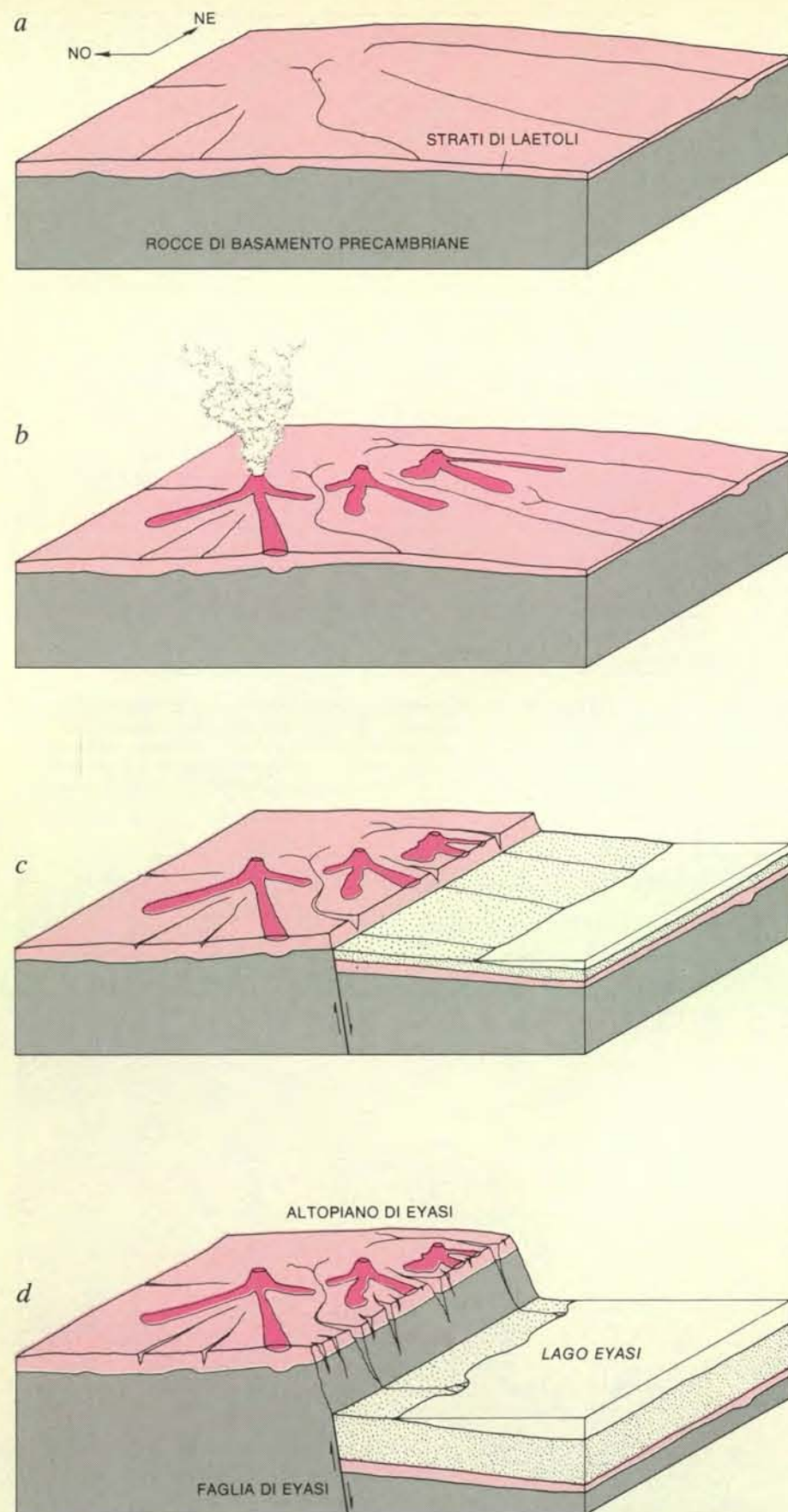
Un altro componente della maggior parte dei tufi formati da ceneri indisturbate, e forse di tutti, è la carbonatite. Questo materiale igneo è formato da carbonato di calcio, con o senza carbonato di sodio. La carbonatite è molto rara. Di un solo vulcano al mondo, l'Oldoinyo Lengai, 90 chilometri a nord di Laetoli, si sa che ha eruttato carbonatite in tempi recenti. La carbonatite dell'Oldoinyo Lengai e la carbonatite degli strati di Laetoli appartengono al tipo ricco di carbonato di sodio, e ciò ci fornisce un indizio per comprendere la conservazione delle impronte nel Tufo delle Orme.

Una carbonatite che contenga sia carbonato di calcio sia carbonato di sodio reagisce con l'acqua in un modo tipico. Il carbonato di calcio diventa un precipitato fine e il carbonato di sodio entra a far parte della soluzione. Quando la soluzione evapora, si formano cristalli di trona, i quali forniscono agli strati di cenere un cemento istantaneo.

Il Tufo delle Orme è uno dei tre tufi indisturbati laminati degli strati di Laetoli. Orme fossili sono state poi trovate in tutt'e tre i tufi stratificati, ma sono particolarmente abbondanti e diffuse nel cosiddetto Tufo delle Orme. Di spessore generalmente compreso fra 12 e 15 centimetri, esso si trova in prossimità del limite superiore della zona fossilifera ed è ricoperto da un più consistente deposito di ceneri, che in taluni punti raggiunge lo spessore di 50 centimetri. I sottili strati di cenere che costituiscono il Tufo delle Orme si accumularono nelle prime fasi di un episodio eruttivo, mentre il tufo più spesso sovrastante rappresenta il culmine dello stesso episodio.

Il Tufo delle Orme è divisibile in due unità principali. Quella inferiore ha uno spessore compreso fra sette e 10 centimetri; quella superiore fra quattro e sei. La divisione fra le due unità è netta e la parte superiore dell'unità più profonda risulta in vari punti erosa. Le due unità si differenziano per i caratteri litologici, per struttura e per le orme che contengono.

L'unità inferiore è suddivisa in 14 strati sottili, il cui spessore varia fra due e 15 millimetri. Gli strati sono generalmente identificabili nell'intera area di Laetoli; ciascuno di essi è, con ogni probabilità, il prodotto di una singola eruzione vulcanica. La maggior parte degli strati ricoprono irregolarità superficiali, come le orme, con piccole variazioni di spessore. Ciò dimostra che essi rimasero essenzialmente indisturbati là dove si depositarono.

La superficie di cinque strati risulta fittamente bucherellata dall'impatto di gocce di pioggia; altri tre presentano le impronte di gocce di pioggia solo in zone

**Il lago Eyasi, a sud-est del lago Vittoria in Tanzania, è prossimo all'estremità sud della Eastern Rift Valley, la grande fossa tettonica che corre attraverso il Kenya e l'Etiopia sino al Mar Rosso. Resti di antichi ominidi sono stati trovati in gran numero nella valle e nelle sue vicinanze. Alcune delle località più note in cui sono stati trovati resti di ominidi sono Hadar e Omo in Etiopia, la zona orientale del lago Turkana in Kenya, la Gola di Olduvai e la vicina località di Laetoli in Tanzania.**

**Gli strati di Laetoli, depositi di cenere vulcanica (che fu eruttata dal vulcano Sadiman) a nord del lago Eyasi, coprono un'area di 1500 chilometri quadrati *(in colore chiaro)*. Essi sono sovrapposti a rocce di basamento precambriane *(in grigio)* a ovest e sono ricoperti da roccia vulcanica *(in colore scuro)* a est. Gli strati più ricchi di fossili si trovano nell'area di 70 chilometri quadrati di Laetoli, con contorno in colore. Il Tufo delle Orme, vicino alla parte superiore degli strati fossiliferi, ha in generale, all'interno della linea ovale nera, uno spessore di più di 12 centimetri.**

**Questi diagrammi illustrano lo sviluppo della regione di Eyasi. All'epoca in cui si depositarono le ceneri degli strati di Laetoli, fra 3,8 e 3,5 milioni di anni or sono *(a)*, la regione era una savana erbosa relativamente piatta. Circa 1,2 milioni di anni dopo *(b)*, essa attraversò un periodo di maggiore attività vulcanica. Dopo altri 1,2 milioni di anni *(c)*, quando già aveva avuto inizio la fagliatura della fossa tettonica, l'altopiano di Eyasi si sollevò e, ai suoi piedi, si verificò la formazione di un lago. Un ulteriore sollevamento e l'erosione *(d)* resero il paesaggio della regione di Eyasi qual è oggi.**

limitate. Le impronte delle gocce d'acqua, molto ravvicinate fra loro e ben definite, furono provocate evidentemente da acquazzoni abbastanza fitti da inumidire la cenere, ma non abbastanza da eroderla. Le orme sono impresse sulla superficie dello strato superiore (il n. 14) e su quella di sette dei tredici strati inferiori. Sono molto abbondanti e diffuse negli strati n. 9 e n. 14; pochissime sono invece quelle all'interno di singoli strati di cenere.

L'unità superiore del Tufo delle Orme presenta un quadro un po' diverso. Essa è formata da quattro soli strati, ciascuno dei quali rappresenta o una singola eruzione vulcanica o una serie di eruzioni susseguitesi a breve intervallo di tempo. Ciascuno strato ha subito un'estesa rideposizione per azione dell'acqua: il deflusso superficiale associato ad abbondanti piogge. I due strati inferiori ricoprono irregolarità della superficie, ma sono entrambi più spessi in corrispondenza di depressioni, come orme di zoccoli animali, di quanto non siano su aree in rilievo, come i bordi delle orme. La maggior parte delle aree in rilievo sono troncate alla base del terzo strato, che in molti punti è una superficie di erosione. Là dove sono stati ridepositati, gli strati presentano una laminazione e il tufo laminato di ciascuno strato può riempire i canali ampi e poco profondi prodottisi negli strati sottostanti. Nell'unità superiore le orme sono assai diffuse sulle superfici degli strati n. 1 e 2. Impronte di pioggia sono state trovate in una zona sulla superficie dello strato n. 2.

Come abbiamo osservato sopra, la maggior parte della cenere nel Tufo delle Orme è formata da particelle di lava contenenti il minerale di calcio melilite. Le particelle, di forma ovoidale o sferoidale, hanno per lo più le dimensioni di granelli di sabbia da fine a media (da 0,1 a 0,3 millimetri). Particelle di calcite grossolanamente cristallina, di forma fra arrotondata e appiattita e di lunghezza media fra 0,02 e 0,01 millimetri, sono diffuse, e localmente abbondanti, nello strato più basso dell'unità inferiore (strato n. 1). Particelle di calcite simili sono state trovate in alcuni luoghi negli strati n. 2-5.

La calcite contiene concentrazioni relativamente alte di stronzio e di bario (un segno tipico della presenza della carbonatite). Elementi strutturali visibili al microscopio dimostrano però che la calcite ha sostituito quello che era in origine un minerale di carbonatite ricco di sodio come la nyererite. Le particelle di calcite rappresentano indubbiamente solo una piccola frazione delle particelle di cenere di carbonatite presenti in origine; la maggior parte della cenere dev'essere scomparsa reagendo con acqua (come avviene, per esempio, nel processo che conduce alla formazione di cristalli di trona).

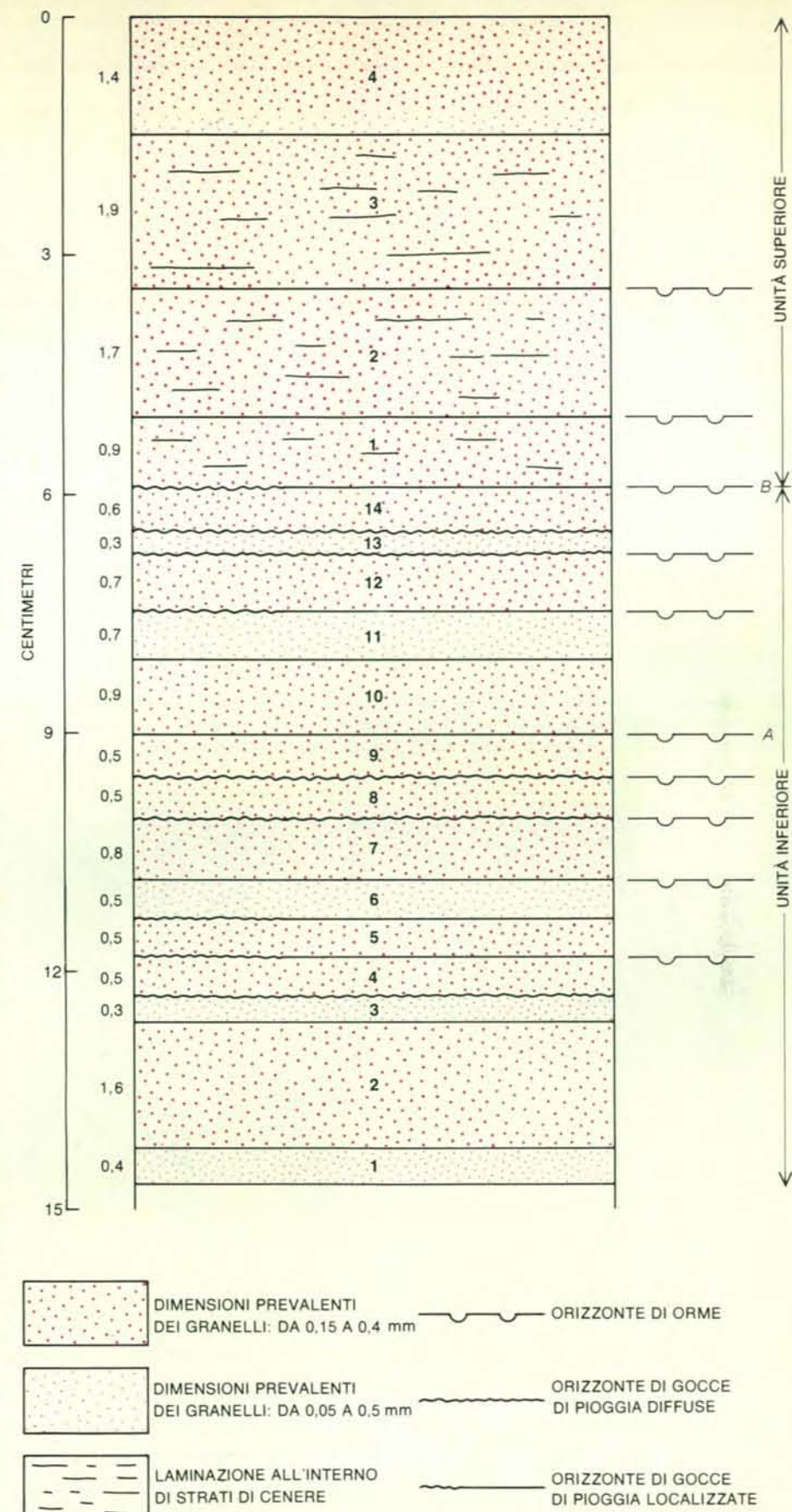
Il trona, essendo un minerale solubile in acqua, si dissolse molto tempo fa; gli strati del Tufo delle Orme sono oggi cementati da calcite e, più di rado, dal minerale phillipsite (un silicato di alluminio). Alcuni strati constano in gran parte di cemento in cui le particelle di cenere sembrano «galleggiare». Questi strati vengono interpretati come depositi di ceneri in cui gran parte dello spazio oggi riempito da cemento era occupato in origine da carbonatite. Gli strati ricchi di cemento sono comuni nell'unità inferiore del tufo e rari nell'unità superiore, la quale ricevette presumibilmente una proporzione minore di cenere di carbonatite.

La superficie del terreno sepolta sotto il Tufo delle Orme era pressoché piatta. Le sue irregolarità maggiori consistevano in depressioni della profondità di circa 10 centimetri, che vengono interpretate come piste attraverso la savana del Pliocene. La presenza di erba è suggerita da piccole radici fossilizzate sotto il Tufo delle Orme. Quando si depositarono i primi strati del Tufo delle Orme, il terreno doveva essere però essenzialmente spoglio, dato che gli steli verticali d'erba avrebbero disturbato la delicata stratificazione della cenere. È molto probabile che l'erba fosse stata completamente brucata dagli erbivori della savana.

Alla base del Tufo delle Orme abbonda sterco fossilizzato di conigli, giraffe e di quell'antilope in miniatura nota come madocca o dig-dig. Vi sono stati trovati anche rami e ramoscelli fossilizzati di alberi e cespugli. Negli strati più alti dell'unità inferiore sono stati trovati anche rametti, spine e foglie di acacia, albero tipico della savana; uno strato contiene sterco di conigli e forse anche di giraffe.

Il Tufo delle Orme fu depositato nell'arco di un breve periodo di tempo, forse in alcune settimane, tra la fine della stagione secca della savana e l'inizio della stagione delle piogge. L'eccellente conservazione delle orme e delle impronte di pioggia dimostra che gli strati di tufo devono essere stati sepolti da ceneri vulcaniche pochissimo tempo dopo la formazione delle impronte. Un ulteriore indizio del fatto che il processo di seppellimento fu rapido è fornito dalla continuità degli strati, alcuni dei quali dello spessore di soli pochi millimetri, su un'area di 70 chilometri quadrati di estensione.

Ma esiste anche un'altra prova della breve durata temporale dell'evento. L'assenza di steli d'erba alla base del tufo suggerisce che le prime eruzioni vulcaniche abbiano avuto luogo nella stagione secca. Le impronte di pioggia nell'unità inferiore furono prodotte probabilmente da brevi rovesci, quando la stagione secca si avviava alla sua conclusione. Mentre la rideposizione delle ceneri nell'unità superiore in conseguenza del deflusso superficiale di acque indica un periodo di piogge più abbondanti, presumibilmente l'inizio della stagione delle piogge. Inoltre la diffusa erosione alla base dell'unità superiore e fra gli strati n. 2 e n. 3 dell'unità superiore è chiaramente attribuibile a piogge abbondanti. Infine la scarsità di impronte di pioggia sulla superficie degli strati dell'unità superiore può essere ricondotta al fatto che i sedimenti saturi di pioggia non conservano tali impronte.

Poiché il Tufo delle Orme è ben cementato e più resistente all'erosione di quanto non siano i depositi sottostanti e sovra-

**Schema del Tufo delle Orme in sezione in un tratto dello spessore di quasi 15 centimetri. Delle 14 suddivisioni dell'unità inferiore, otto presentano impronte particolarmente abbondanti negli orizzonti *A* e *B*. Anche due suddivisioni dell'unità superiore presentano impronte. Le orme di ominidi appaiono nell'orizzonte *B*. I numeri indicano lo spessore medio degli strati nel Sito *A*.**

stanti, esso è sopravvissuto all'erosione pur essendo rimasto esposto in aree piuttosto estese. Là dove è eroso dagli agenti atmosferici, il tufo si spacca facilmente lungo i piani di divisione degli strati, rivelando tutte le orme presenti sulla superficie dello strato esposto. A Laetoli sono state finora esaminate le orme in 16 siti; i siti esposti sono stati designati con le lettere dell'alfabeto inglese da *A* a *P*.

La più estesa fra le aree esposte è il Sito *A*, che conserva tracce evidenti su un'area di circa 800 metri quadrati. Per pervenire a una stima della densità delle impronte, abbiamo contato il numero di impronte su 17 metri quadrati del Sito *A* e 20 metri quadrati del Sito *D* (che copre circa 113 metri quadrati). Le piccole impronte ovali di conigli (e forse di madocche) ammontavano a una media di 21 per metro quadrato nel Sito *A* e di 30 per metro quadrato nel Sito *D*. Altre impronte, per lo più di dimensioni maggiori, ammontavano a una media, rispettivamente, di 1,8 e 3,6 per metro quadrato nei due siti. Calcolando una media di 23 orme di varie dimensioni per metro quadrato, nel solo Sito *A* sarebbero conservate 18 400 impronte.

Sono state identificate le impronte di animali appartenenti a un totale di 17 famiglie diverse. Questi animali rappresentano la maggior parte di quelli di cui sono stati trovati i resti fossili negli strati di Laetoli. La distribuzione verticale delle impronte negli strati si accorda con l'ipotesi che esse siano state formate nel periodo di transizione fra la stagione secca e quella delle piogge. Perciò le impronte presenti nell'unità inferiore sono principalmente quelle di conigli, di galline faraone, di rinoceronti e di altri animali che vivono nelle savane erbose durante la stagione secca. Le impronte nell'unità superiore comprendono invece quelle degli abitatori della stagione umida: proboscidati, cavalli, babbuini e ominidi. Nessuna loro impronta si trova nell'unità inferiore. Anche le impronte di antilopi di taglia maggiore sono più comuni nell'unità superiore che in quella inferiore. Esse documentano con ogni probabilità le migrazioni delle stagioni umide tipiche della savana dell'Africa Orientale. Le tracce seguono le piste preesistenti e indicano una normale andatura di trasferimento. Da ciò possiamo concludere che le piogge di ceneri del vulcano Sadiman non furono così abbondanti da modificare il normale itinerario delle migrazioni stagionali.

Le cinque orme lasciate probabilmente da ominidi, scoperte nel 1977, si trovano nel Sito *A*. Le due sequenze parallele di orme di ominidi scoperte nel 1978 si trovano nel Sito *G*. Le due sequenze, distanziate fra loro di circa 25 centimetri, sono state riportate in luce su una lunghezza di 25 metri. Le orme più nitide sono profonde da uno a 3 centimetri e hanno margini netti. Esse rivelano il tallone arrotondato, l'arco ben pronunciato e l'alluce rivolto in avanti tipici del piede umano. Una serie di orme fu prodotta da un solo individuo di piccola statura; l'altra è composita: le orme originarie furono lasciate da un individuo piuttosto grande e, sulla prima serie di impronte, se ne sovrappone una seconda lasciata da un individuo più piccolo che camminava ponendo il piede sulle orme originarie.

Le orme di ominidi ci consentono di concludere che 3,5 milioni di anni or sono questi precursori africani dell'uomo antico avevano già un'andatura bipede pienamente eretta. In tale periodo gli ominidi africani erano ancora piccoli, rispetto alla media degli uomini posteriori, sia per statura sia per volume cerebrale. Supponendo che, come nelle popolazioni umane moderne, la lunghezza dei piedi degli ominidi di Laetoli fosse press'a poco pari al 15 per cento della loro statura, il più piccolo dei tre che lasciarono le loro orme nel sito *G* doveva essere alto 120 centimetri, mentre la statura del secondo doveva essere di circa 140 centimetri. La lunghezza delle orme lasciate dal terzo ominide, il più alto, non può essere misurata perché sono in parte cancellate da quelle del secondo in ordine di statura.

La precoce comparsa della stazione eretta nel corso dell'evoluzione dell'uomo riveste una grande importanza. Essa infatti rese libere le mani sia per il trasporto sia per la produzione e l'uso di utensili. Nonostante le accurate ricerche, negli strati di Laetoli non sono stati rinvenuti utensili litici. Pare perciò probabile che gli ominidi che lasciarono le loro tracce nel Tufo delle Orme non fossero ancora pervenuti alla fase della fabbricazione di utensili di pietra. Rimane comunque il fatto che la stazione eretta consentì loro un uso continuo del primo fra tutti gli utensili dei primati: le mani libere.

Le orme di animali non ominidi conservate nel Tufo delle Orme presentano una vasta gamma di dimensioni. Fra le più grandi sono quelle di elefanti, dinoteri, giraffe e rinoceronti. Fra le più piccole, quelle di un piccolo felino, conigli, galline, faraone, francolini (uccelli di taglia un po' minore della gallina faraona) e un insetto (forse uno scarabeo stercorario).

Tanto la produzione quanto la conservazione di orme fossili richiedono un insieme di circostanze piuttosto particolare. Innanzitutto il terreno dev'essere abbastanza soffice e dotato di una coesione sufficiente a conservare impronte chiare. In secondo luogo le impronte devono venire sepolte prima che processi di degradamento come l'erosione possano modificarle o cancellarle. Infine, il materiale che le ricopre deve in seguito potersi separare in modo netto dalla superficie su cui si trovano le impronte. La maggior parte delle impronte fossili ben conservate furono lasciate in sedimenti fangosi che, asciugando, si indurirono e furono poi ricoperte da sabbia o fango. La cenere vulcanica del Tufo delle Orme era invece del tutto diversa; essa aveva la struttura di granelli di sabbia di dimensioni da

**La grande depressione circolare, poco profonda, in quest'esposizione dell'unità superiore n. 1, un orizzonte proprio della stagione umida, è un'orma di elefante. Nella grande sono visibili altre impronte, probabilmente di animali di piccola taglia. La scala graduata è lunga 20 centimetri.**



**In questa parte del tufo esposto sono visibili alcune orme: a sinistra in basso appaiono le orme di un uccello; a sinistra e a destra, in alto, quelle di un animale di piccole dimensioni, forse un coniglio o una madocca. Vicino al centro, un'impronta di zampa posteriore di babbuino.**

fini a medie. Per la conservazione delle impronte, questo tipo di superficie è estremamente insolito.

La natura delle orme prodotte in una sabbia non consolidata dipende in buona parte dal suo grado di umidità. La sabbia asciutta, incoerente, conserverà abbastanza bene le tracce di animali piccoli, ma non conserverà in modo chiaro le orme profonde di animali di grossa taglia perché le ripide pareti laterali delle impronte franeranno verso l'interno. La sabbia bagnata conserverà le impronte di animali di grandi dimensioni ma non le tracce di piccoli animali. Per conservare entrambi i tipi di tracce, la sabbia dev'essere solo leggermente umida.

La nitidezza delle tracce di piccoli animali nel Tufo delle Orme dimostra che, quando furono lasciate, la cenere era soffice. Al tempo stesso i margini verticali molto netti di alcune impronte di grossi animali dimostrano che la cenere aveva una buona coesione. (Alcune delle impronte più grandi hanno i bordi lievemente inclinati verso l'interno, la cenere doveva avere quindi un minor grado di compattezza). La maggior parte delle impronte più grandi hanno anche un bordo rialzato, prodotto dalla pressione laterale esercitata dalla zampa. Alcuni centimetri più sotto gli strati di cenere dovevano essere già solidi, perché le orme più grandi, quelle dei rinoceronti e dei proboscidati, hanno una profondità che in genere non supera i cinque centimetri.

È difficile immaginare che quando le varie impronte furono prodotte tutt'e dieci gli strati di cenere che ne hanno conservato le tracce fossero umidi al punto giusto. Ma ciò sarebbe stato strettamente necessario? La risposta è: no. Esattamente come, mescolando cenere di legna o cemento portland con sabbia asciutta, si conferisce alla sabbia maggiore coesione, così la carbonatite potrebbe aver aumentato la coesione della cenere vulcanica, fornendo una matrice a grana fine alle particelle di cenere. Se la carbonatite fosse stata inizialmente a grana fine, sarebbe stata disponibile una matrice primaria idonea proprio al momento della deposizione, e la successiva reazione della carbonatite con acqua avrebbe fornito una matrice secondaria ancora più fine di carbonato di calcio precipitato. Nel frattempo, a meno che il carbonato di sodio disciolto dalla carbonatite non fosse stato completamente dilavato dalla pioggia, all'evaporare dell'acqua si sarebbe cristallizzato in trona, cementando leggermente lo strato superficiale di cenere e fissando le impronte finché non fossero state sepolte dalla pioggia di cenere.

Nelle successive migliaia di anni gli strati di cenere del Tufo delle Orme furono modificati da una serie di processi. Innanzitutto la reazione della carbonatite con acqua produsse un'abbondanza di calcite che, nelle prime fasi, deve aver debolmente cementato gli strati di cenere e deve aver sostituito nella cenere lo sterco e il materiale vegetale. Durante questa fase insetti e altri animali scavavano nella cenere. Per l'azione sui tufi degli agenti atmosferici, cenere e melilite furono trasformate in argilla di montmorillonite e il minerale di silicato di alluminio phillipsite, depositato localmente, funzionò come agente cementante aggiuntivo. Successivamente la calcite riempì quasi tutti i pori restanti, trasformando gli strati di cenere in una roccia dura ben cementata; si depositò anche fra alcuni strati di cenere sotto forma di fogli sottili, discontinui, di calcare. Tale calcare rivestì, nel Sito *G*, molte delle impronte di ominidi facilitandone lo scavo senza correre il rischio di alterarle.

In sintesi, il Tufo delle Orme fu il prodotto di una combinazione di condizioni estremamente insolita. Press'a poco 3,5 milioni di anni or sono l'area di Laetoli ospitava un'abbondante e differenziata popolazione animale. Le piogge di cenere vulcanica che caddero su quest'area non furono abbastanza copiose da indurre gli animali ad andarsene. L'eccellente definizione di una moltitudine di impronte, orme e tracce prodotte nella cenere fresca fu il risultato della mescolanza di carbonatite, un materiale igneo non comune, con le particelle più tipiche della polvere vulcanica; gli strati di cenere che recavano impronte furono poi sepolti a intervalli ravvicinati da nuove piogge di cenere. Infine una pioggia di cenere particolarmente abbondante seppellì il Tufo delle Orme proteggendolo dall'erosione e gli strati contenenti le impronte furono cementati e divennero dura roccia. Quando, recentemente, l'azione degli agenti atmosferici ha spaccato gli strati di tufo lungo i loro piani di stratificazione, una finestra, si può dire unica, si è aperta sul mondo degli antichi ominidi e degli animali che ne condividevano l'habitat.

# Titano

*Il più grande fra i satelliti di Saturno è l'unico nel sistema solare a possedere un'atmosfera consistente, la cui chimica potrebbe essere simile a quella dell'atmosfera terrestre prima dell'origine della vita*

di Tobias Owen

Il 12 novembre 1980 la sonda *Voyager 1* passò a meno di 7000 chilometri da Titano, il più grande fra i satelliti di Saturno. Fu l'incontro più ravvicinato fra un veicolo Voyager e un pianeta o un satellite, ma quest'impresa comportò un costo elevato. Facendo passare *Voyager 1* così vicino a Titano, si rinunciò infatti alla possibilità che la forza gravitazionale di Saturno lo dirigesse in modo da continuare il suo viaggio nel sistema solare verso Urano e Nettuno. Il sacrificio parve però giustificato, in quanto Titano è l'unico satellite del sistema solare in possesso, a quanto si sa, di un'atmosfera consistente. Inoltre il suo colore rossastro, che è una sua prerogativa unica fra i satelliti di Saturno, faceva pensare che la chimica dell'atmosfera di Titano potesse produrre composti colorati. Grazie alla vicinanza alla superficie di Titano, gli strumenti a bordo del veicolo spaziale avrebbero potuto fornire prestazioni ottimali.

I risultati ottenuti hanno dimostrato che la scelta era saggia. Si è accertato infatti che Titano è l'unico corpo noto nel sistema solare, oltre alla Terra, la cui superficie sia coperta almeno in parte da liquidi. Su Titano il liquido è metano. Gli strumenti del Voyager hanno dimostrato inoltre che l'atmosfera di Titano è più densa dell'atmosfera terrestre. Quest'atmosfera ha conservato condizioni molto simili a quelle esistenti forse su tutti i pianeti subito dopo la loro formazione. Specificamente, l'atmosfera di Titano contiene carbonio, azoto e idrogeno, ma è priva di ossigeno molecolare. In queste condizioni le reazioni chimiche che hanno luogo oggi nell'atmosfera di Titano potrebbero dare origine ad alcune fra le molecole organiche che si pensa abbiano precorso l'emergere della vita sulla Terra.

*La presenza di un'atmosfera*

Titano fu scoperto da Christiaan Huygens nella primavera del 1655, lo stesso anno in cui egli suggerì che Saturno doveva essere circondato da anelli. Il satellite ricevette il suo attuale nome quasi due secoli dopo, quando Sir John Herschel assegnò nomi ai sette satelliti di Saturno noti a quel tempo. Il nome Titano fu una scelta felice. Herschel sapeva solo che Titano era il satellite più luminoso di Saturno; dopo di allora è stato dimostrato che è anche il più grande, più grande addirittura del pianeta Mercurio. Per qualche tempo si pensò che fosse anche il satellite più grande di tutto il sistema solare, ma misurazioni eseguite da *Voyager 1* dimostrano che questo primato non gli appartiene. Le misurazioni anteriori erano state gonfiate dallo spessore dell'atmosfera di Titano. Il corpo solido di Titano ha un raggio di 2575 chilometri. Più grande di Titano è il satellite galileiano di Giove Ganimede, che ha un raggio di 2640 chilometri.

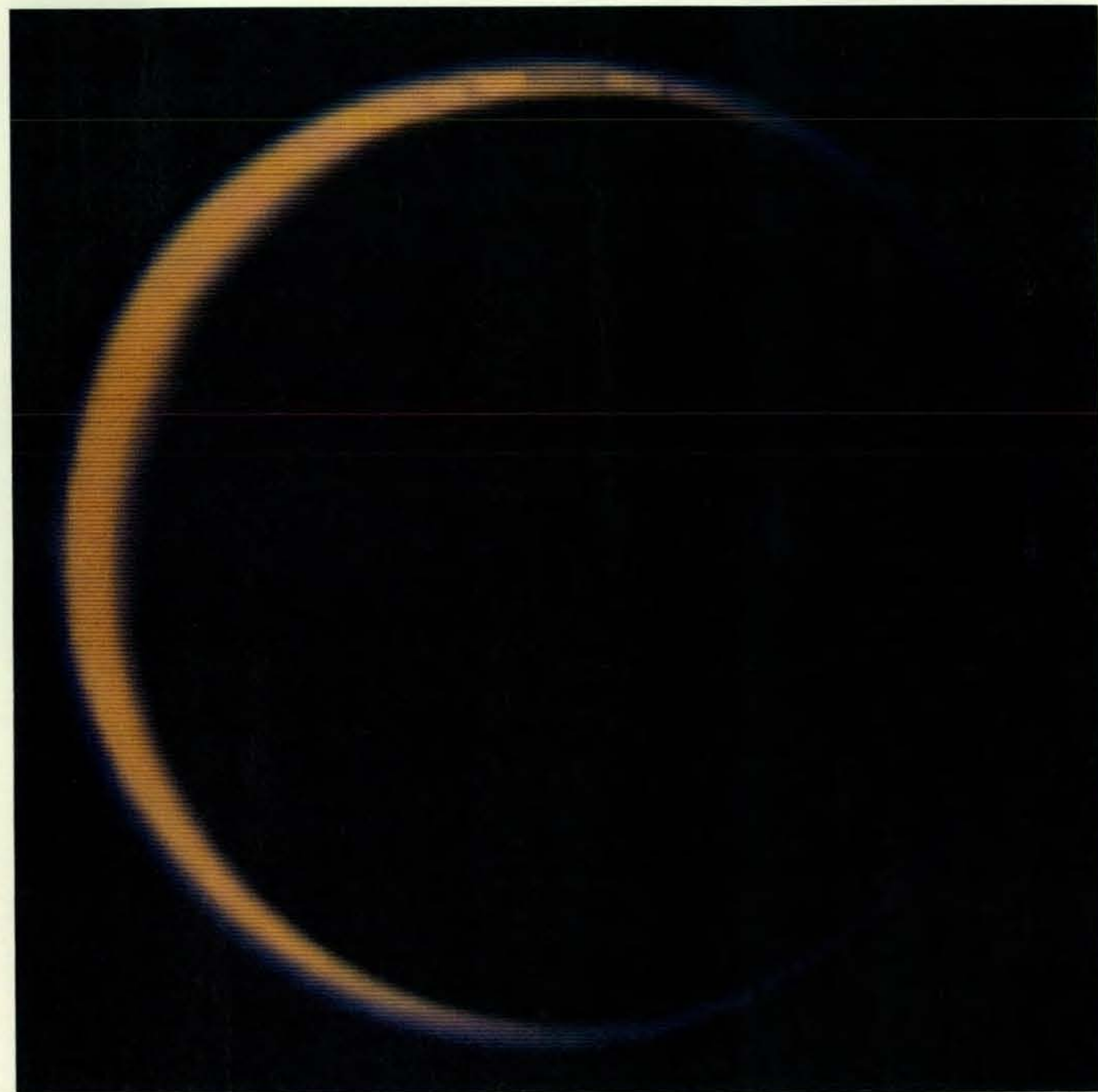
Il primo indizio dell'esistenza di un'atmosfera su Titano risale a osservazioni pubblicate nel 1908 dall'astronomo catalano José Comas Solá. Solá riferì che il minuscolo disco di Titano, quale gli appariva al telescopio, presentava un oscuramento al bordo, fenomeno di cui egli suggerì come causa il possesso di un'atmosfera: la luce solare riflessa verso la Terra dalla periferia del disco di Titano doveva passare attraverso una parte dell'atmosfera del satellite maggiore di quella attraversata dalla luce solare riflessa dalla parte centrale del disco. La luce proveniente dal bordo risulterebbe quindi maggiormente indebolita a causa dell'assorbimento nell'atmosfera di Titano.

È difficile stabilire se Solá abbia osservato realmente un oscuramento al bordo sul disco di Titano. Oggi sappiamo che le descrizioni da lui date di nubi a chiazze sui satelliti maggiori di Giove sono erronee. Pare nondimeno che le sue osservazioni abbiano indotto Sir James Jeans a includere Titano e i satelliti giganti di Giove nel suo studio teorico sull'allontanamento dei gas atmosferici dai corpi del sistema solare. Nel 1916 Jeans concluse che Titano ha conservato probabilmente un'atmosfera, grazie alla sua bassa temperatura, nonostante la relativa esiguità del suo volume e della sua attrazione gravitazionale rispetto, per esempio, alla Terra. La distanza di Titano dal Sole, combinata con una stima ragionevole della sua riflettività (e conseguente perdita di calore solare), conducono a prevedere per la sua superficie e la sua atmosfera, una temperatura compresa fra 60 e 100 kelvin. Per un tale intervallo di temperature l'opera di Jeans dimostra che una sostanza gassosa il cui peso molecolare sia di 16 o più non sarebbe sfuggita da Titano nel corso dell'intera storia del sistema solare.

Varie sostanze soddisfano il limite di Jeans sul peso. Una di esse è l'ammoniaca ($NH_3$), il cui peso molecolare è 17. Negli anni trenta Rupert Wildt, dell'Università di Göttingen, la identificò come un componente dell'atmosfera di Giove. Wildt aveva trovato che lo spettro della luce solare riflessa da Giove nell'infrarosso dimostrava l'assorbimento della radiazione solare alle lunghezze d'onda tipiche delle molecole di ammoniaca. Usando un metodo simile, Theodore Dunham, Jr., dell'Osservatorio di Mount Wilson scoprì la presenza di ammoniaca su Saturno. Alla temperatura ipotizzata per Titano, però, l'ammoniaca avrebbe dovuto trovarsi allo stato solido e non poteva quindi costituire un componente di rilievo dell'atmosfera. Altre sostanze che soddisfano il limite di Jeans sono l'argo, il neo e l'azoto molecolare ($N_2$): tutte dovevano essere presenti con una concentrazione apprezzabile nel miscuglio di gas e di polvere che si condensò a formare il sistema solare. Il problema è che è difficile rivelarli spettroscopicamente. Nessuno di essi, infatti, assorbe molta radiazione nell'infrarosso.

Un'altra sostanza ancora che soddisfa quel limite è il metano ($CH_4$), il cui peso molecolare è 16. A differenza dell'argo, del neo e dell'azoto molecolare, esso presenta una serie marcata di bande di assorbimento nell'infrarosso e a differenza dell'ammoniaca è gassoso alla temperatura prevista per Titano. Nel 1932 Wildt identificò il metano negli spettri di Giove, Saturno, Urano e Nettuno. Poi, nel 1944, Gerard P. Kuiper dell'Università di Chicago ne accertò la presenza nello spettro di Titano. Questa scoperta costituì la prima prova incontrovertibile del possesso di un'atmosfera da parte del massimo satellite di Saturno. Confrontando lo spettro di Titano con spettri di laboratorio del metano a basse pressioni, Kuiper ne dedusse che l'assorbimento della luce del Sole ad opera del gas lungo un itinerario verticale nell'atmosfera di Titano è equivalente all'assorbimento di tale radiazione per opera di una colonna di metano lunga 200 metri alla pressione di un'atmosfera e a una temperatura di 273 kelvin. (Tali condizioni sono chiamate temperatura e pressione standard, TPS.) Ai fini del confronto, una colonna verticale attraverso l'atmosfera terrestre equivale a una colonna lunga 8 chilometri a TPS.

Alcuni problemi connessi a quelle prime nozioni sull'atmosfera di Titano cominciarono a prender forma nel 1965, quando Frank J. Low dell'Università dell'Arizona dedusse dalla luminosità della radiazione emessa da Titano alla lunghezza d'onda di 10 micrometri, nell'infrarosso, che la temperatura del satellite è di 165 kelvin, una temperatura quasi doppia di quella imputabile al semplice riscaldamento solare della superficie e dell'atmosfera inferiore di Titano. Le scoperte di Low passarono pressoché inosservate per sette anni, dopo di che un certo numero di ricercatori cominciarono a imbattersi in altre sorprese. Da un lato, i calcoli della temperatura di Titano fondati su misurazioni della sua luminosità a varie lunghezze d'onda nell'infrarosso non riuscivano a dare un valore coerente, ma anche la «luminosità» di Titano nelle lunghezze d'onda radio forniva contraddizioni. La luminosità radio in un'osservazione implicava una temperatura alla superficie di 200 kelvin.

Inoltre la luce riflessa da Titano ai piccoli angoli che si formano per lo più fra il Sole, Titano e la Terra risultò avere una polarizzazione positiva: il vettore che rappresentava l'intensità massima della

**L'atmosfera di Titano è visibile in quest'immagine della faccia notturna del satellite ripresa il 25 agosto 1981 dal veicolo spaziale *Voyager 2*. La falce arancione che sagoma la parte sinistra del bordo di Titano è formata dalla luce del Sole riflessa da un aerosol: uno strato di particelle solide sospese nell'atmosfera di Titano 200 chilometri circa al di sopra della superficie. L'alone azzurro che circonda l'intero bordo di Titano rappresenta la diffusione della luce solare a grandi angoli ad opera di particelle di nebbia sospese a ben 300 chilometri al di sopra della superficie.**

componente elettrica del campo elettromagnetico della luce era perpendicolare al piano dell'angolo. Questa scoperta, fatta indipendentemente da Joseph Veverka della Cornell University e da Benjamin H. Zellner dell'Università dell'Arizona, suggeriva che la luce riflessa da Titano provenisse non dalla superficie solida del corpo dopo avere attraversato un'atmosfera trasparente, ma da un'atmosfera profonda e densa di nubi. Marte, per esempio, ha solo un'atmosfera molto rarefatta e la luce che riflette verso la Terra a piccoli angoli presenta una polarizzazione negativa; il massimo del vettore elettrico è nel piano in cui si trova l'angolo.

Nel frattempo Laurence M. Trafton dell'Università del Texas ad Austin aveva ipotizzato che o la quantità di metano presente su Titano era molto maggiore di quella dedotta da Kuiper, oppure su Titano doveva essere presente in abbondanza anche qualche altro gas. Con l'aiuto di un intensificatore d'immagine nell'infrarosso che era appena divenuto disponibile, Trafton aveva trovato nello spettro di Titano una banda di assorbimento del metano, in prossimità della lunghezza d'onda di un micrometro, inaspettatamente marcata. L'intensità della banda poteva essere dovuta alla presenza di metano in grande quantità, oppure a collisioni fra molecole di metano e molecole di un gas non rivelato, che perturbassero gli stati di vibrazione delle molecole di metano, facendo sì che queste assorbissero radiazione infrarossa su un ampio intervallo di lunghezze d'onda. In entrambi i casi l'intensità della banda di assorbimento doveva essere in relazione sia con l'abbondanza del metano sia con la pressione atmosferica locale.

**La faccia illuminata di Titano è stata fotografata da *Voyager 1* il 9 novembre 1980, tre giorni prima di avvicinarsi a una distanza di meno di 7000 chilometri dal satellite. In questa immagine la superficie di Titano è occultata dall'ininterrotta opacità dello strato di aerosol. L'emisfero sud è più brillante dell'emisfero nord, forse in conseguenza di un mutamento stagionale nel tasso di produzione delle particelle di aerosol. Nell'emisfero settentrionale è primavera; l'emisfero meridionale è quindi appena passato per un'estate della durata di ben sette anni.**

Trafton scoprì inoltre che le più intense fra le bande di assorbimento del metano nello spettro infrarosso di Titano hanno un aspetto diverso da quello delle stesse bande negli spettri di Giove e di Saturno. Le bande di Titano sono meno profonde (meno intense) e più larghe. In base a questa scoperta, prese in considerazione l'idea della presenza, nell'atmosfera di Titano, di piccole particelle in quantità e ad altitudini diverse rispetto alle particelle che compongono le nebbie osservate su Giove e su Saturno. Le particelle diffonderebbero la luce del Sole verso la Terra, rendendo in tal modo più intense le bande di assorbimento. Esse inoltre diffonderebbero la luce del Sole a certi angoli nell'atmosfera di Titano. L'allungamento del percorso di un fotone (un quanto di luce) diffuso in questo modo aumenta la probabilità che esso incontri una molecola di metano e ne venga assorbito. La probabilità aumenta anche per fotoni la cui lunghezza d'onda sia un po' diversa da quella che definisce il centro di una banda di assorbimento. (Secondo la teoria quantistica, tali assorbimenti sono improbabili ma non impossibili.) In questo modo le bande vengono allargate.

**L'immagine di un satellite o di un pianeta in possesso di un'atmosfera *(in alto)* è più scura al bordo, cioè alla periferia, che al centro. L'oscuramento è dovuto al fatto che la luce riflessa dal bordo attraversa uno spessore maggiore dell'atmosfera rispetto alla luce riflessa dal centro. Essa viene perciò indebolita in maggior misura dall'assorbimento atmosferico. Per contrasto l'immagine di un pianeta o di un satellite privo di atmosfera *(in basso)* è un disco di luminosità più o meno uniforme. Nel 1908 l'astronomo catalano José Comas Solá ipotizzò che Titano dovesse avere un'atmosfera, dato che la sua immagine al telescopio presentava un oscuramento al bordo.**

*Modelli per Titano*

All'Università statale di New York a Stony Brook, Barry L. Lutz, Robert D. Cess e io esaminammo una lunga serie di spettri di laboratorio del metano eseguiti da Lutz, nel tentativo di migliorare la nostra comprensione dei pianeti esterni e di Titano. Trovammo che le bande di assorbimento studiate da Wildt e da Kuiper sono insensibili alla pressione. Ciascuna banda è formata da strette righe di assorbimento intervallate in modo che qualsiasi allargamento delle singole righe in conseguenza di un aumento della pressione del metano resta senza effetto sulla banda complessiva. Dall'intensità delle bande deducemmo l'abbondanza del metano su Titano. La sua quantità era equivalente a una colonna di metano lunga circa 120 metri a TPS. Poi, dalla relazione fra abbondanza e pressione proposta da Trafton deducemmo la pressione atmosferica in grado di allargare le bande del metano su Titano. (In quest'analisi ignorammo gli effetti della luce diffusa da particelle.)

I risultati furono una sorpresa. Trovammo che la pressione dell'atmosfera di Titano alla base della zona dell'atmosfera che riflette la luce è di circa 400 millibar, ossia pari a quasi metà della pressione atmosferica a livello del mare sulla Terra, la quale è press'a poco di un bar. Il metano di per sé avrebbe potuto dare origine a una pressione di solo un millibar. La spiegazione sembrava risiedere nella seconda delle soluzioni alternative proposte da Trafton, con la presenza su Titano, in grande quantità, di un gas diverso dal metano.

Mentre tentavamo di dedurre la pressione dell'atmosfera di Titano, altri ricercatori si sforzavano di trovare una spiegazione alle misure contraddittorie della temperatura del satellite. La prima idea era stata che la temperatura apparente relativamente alta di Titano, come i 165 kelvin riferiti da Low, potesse essere il risultato di un «effetto serra». Questo concetto presuppone che l'atmosfera di Titano sia trasparente alla luce visibile, ma opaca per l'infrarosso, ossia che la luce del Sole penetri sino alla superficie e la riscaldi e che poi l'atmosfera inferiore intrappoli la radiazione infrarossa riemessa dalla superficie. L'accumularsi di misure contraddittorie aveva compromesso quest'idea, lasciando supporre che l'atmosfera avesse una struttura insolita, con gli strati superiori più caldi della superficie del satellite. Di conseguenza la nozione dell'effetto serra era stata soppiantata da un modello presentato da Robert E. Danielson e da John J. Caldwell della Princeton University.

In tale modello la luminosità di Titano nell'infrarosso è attribuita all'emissione di radiazione infrarossa per opera di un'inversione nell'atmosfera: uno strato in cui la temperatura dell'atmosfera aumenta con l'altitudine anziché diminuire. Specificamente, in tale modello l'atmosfera si riscalda in conseguenza del fatto che la luce solare è assorbita da particelle componenti un aerosol (una sospensione di particelle) nella parte alta dell'atmosfera. La radiazione infrarossa rilevata dalla Terra è emessa dal metano e anche da gas come l'etano ($C_2H_6$) e l'acetilene ($C_2H_2$), prodotti di reazioni chimiche indotte dalla luce in cui molecole di metano sono distrutte dalla radiazione ultravioletta del Sole. Secondo una successiva elaborazione del modello, ad opera di Caldwell, l'atmosfera di Titano sarebbe composta per il 90 per cento da metano. La sua pressione alla superficie è di 20 millibar e la sua temperatura alla superficie è di 86 kelvin, un valore in accordo con la previsione che si potrebbe fare sulla base del semplice riscaldamento solare di un corpo alla distanza di Titano dal Sole. Gli indizi a favore di una pressione alla superficie maggiore di quella contemplata da Caldwell potrebbero essere rifiutati sulla base del fatto che la diffusione della luce solare ad opera delle particelle di aerosol può dare alle bande di assorbimento del metano lo stesso aspetto che assumerebbero per la pressione di un gas diverso dal metano.

Un'alternativa a questo modello è stata proposta da Donald M. Hunten dell'Uni-

**In questa fotografia del polo sud di Titano eseguita da *Voyager 2* si distingue uno strato formato da particelle di nebbia a grande altitudine. Lo strato di nebbia è concentrico alla forma luminosa, che è la punta della falce formata dallo strato opaco di aerosol di Titano illuminato dal Sole.**

versità dell'Arizona. Hunten ha messo in rilievo che la dissociazione di molecole di ammoniaca indotta dalla luce nell'atmosfera di Titano potrebbe condurre all'accumulo di azoto molecolare; l'idrogeno liberato dalla dissociazione sfuggirebbe rapidamente dal satellite. Poiché l'azoto è trasparente alla luce visibile e alla radiazione infrarossa, non potrebbe essere rilevato spettroscopicamente dalla Terra. Hunten dimostrò però che se l'atmosfera di Titano dovesse comprendere abbastanza azoto molecolare da contribuire con 20 bar alla pressione alla superficie, la maggiore pressione potrebbe condurre a un maggiore assorbimento di radiazione infrarossa ad opera dell'azoto stesso, attraverso assorbimento indotto da collisioni. In questo modo un effetto serra potrebbe riscaldare la superficie di Titano sino alla temperatura di 200 kelvin dedotta da una delle misurazioni radio esistenti.

Il modello di Hunten comprendeva lo strato di aerosol e la corrispondente inversione termica che il modello di Danielson e di Caldwell aveva collocato nell'atmosfera superiore. Al di là di questo, però, i modelli erano in stridente contrasto. Per sintetizzare, Danielson e Caldwell ipotizzavano che la temperatura alla superficie di Titano fosse di 86 kelvin e che l'atmosfera contenesse un 90 per cento di metano; Hunten ipotizzava una temperatura alla superficie di 200 kelvin e un'atmosfera composta al 90 per cento da azoto. La discordanza più grave riguardava la pressione: Danielson e Caldwell ipotizzavano una pressione alla superficie di 20 millibar, mentre Hunten proponeva 20 bar, un valore ben mille volte più grande!

Sembrava possibile che questa gravissima discrepanza potesse essere ridotta, e forse addirittura eliminata, da misurazioni migliori della luminosità di Titano in lunghezze d'onda radio. L'utilità di tali misurazioni consisteva nel fatto che la maggior parte dei citati componenti di un'atmosfera nel sistema solare esterno, compresi il metano e l'azoto, sono trasparenti a lunghezze d'onda radio; le misurazioni perciò avrebbero rappresentato probabilmente l'emissione termica dalla superficie di Titano e non da uno strato dell'atmosfera. Un problema che emerse in tali misurazioni è che la debole emissione da Titano viene sommersa dalla radiazione emessa da Saturno. Per venire a capo del problema, Walter J. Jaffe del National Radio Astronomy Observatory, lavorando con Caldwell e con me, si servì del Very Large Array di radiotelescopi nel New Mexico. Facemmo osservazioni di Titano alle lunghezze d'onda radio di 1,3, 2 e 6 centimetri. Il Very Large Array ha una risoluzione angolare sufficiente per separare l'emissione di Titano da quella di Saturno. Trovammo che Titano ha una temperatura alla superficie di 87 kelvin più o meno nove gradi. La nostra scoperta si accordava con il modello di Danielson e di Caldwell. Essa poteva conciliarsi anche con il modello di Hunten, modificato però nel senso di ridurre a un massimo di 2 bar il contributo dell'azoto alla pressione alla superficie. Avevamo così ridotto la discrepanza, ma solo di un fattore 10.

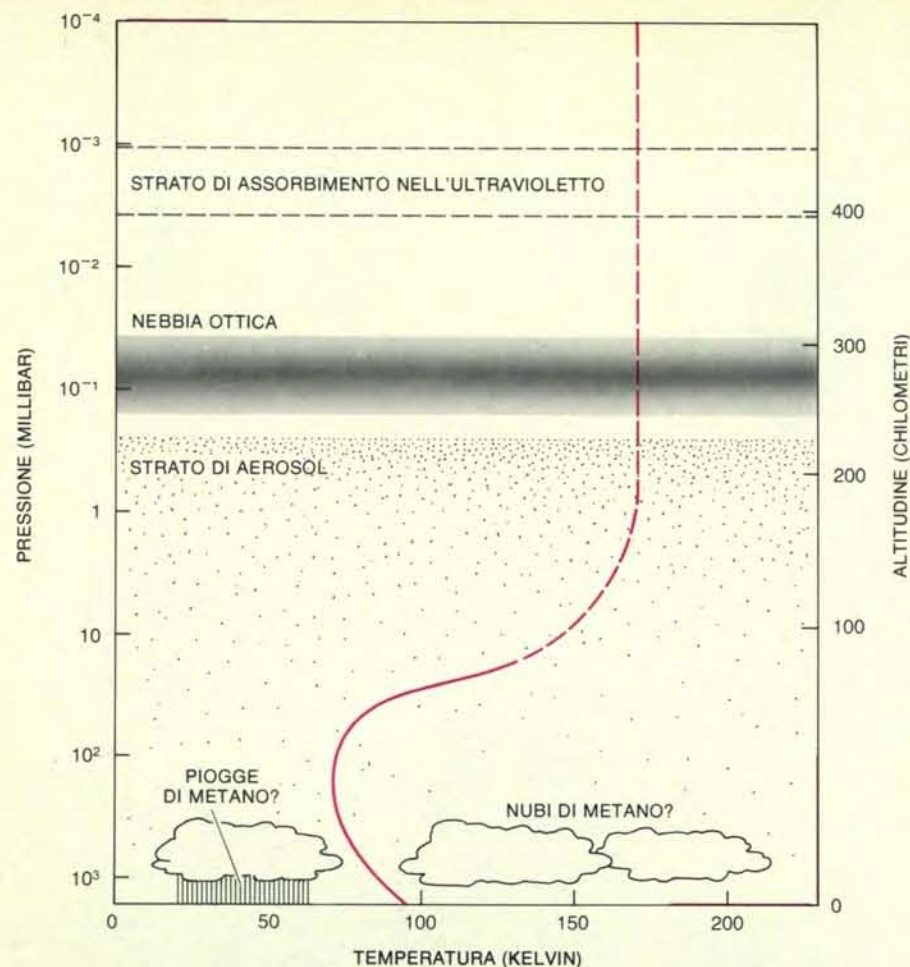
*L'esplorazione con i Voyager*

Era questa la situazione nell'autunno del 1980, quando il *Voyager 1* si approssimò al sistema di Saturno. Si sapeva che Titano era un grande satellite con un'atmosfera almeno tre volte più densa di quella di Marte (dove la pressione media alla superficie è di 7 millibar). La presenza ipotizzata di etano, di acetilene e di un aerosol ad alta quota lasciava supporre un'attiva fotochimica. Il metano era stato rilevato spettroscopicamente; la presenza di azoto era stata proposta. Thomas W. Scattergood e io formulammo l'ipotesi che il colore rosso di Titano fosse una prova indiretta della presenza di azoto. Scattergood aveva tentato di ottenere composti colorati bombardando vari miscugli di gas con protoni ad alta energia. Il suo intento era quello di simulare il bombardamento dell'atmosfera di Titano ad opera di particelle subatomiche elettricamente cariche intrappolate nel campo magnetico che circonda Saturno. I suoi sforzi non dettero alcun risultato negli esperimenti in cui il gas preso in considerazione era solo metano. Se invece esso era costituito da un miscuglio di metano e azoto, il bombardamento conduceva alla formazione di un materiale bruno-rossastro. Ma quanto azoto poteva essere presente su Titano? E quali altri gas c'erano?

I primi risultati dell'arrivo di *Voyager 1* in prossimità di Titano furono immagini, in verità piuttosto deludenti, del corpo del satellite. Alcuni ricercatori avevano sperato di scorgere vuoti nello strato di aerosol, per poter gettare uno sguardo sulla superficie di Titano. Le immagini presentavano invece un satellite che sembrava una palla da tennis lanuginosa e senza cuciture. L'aerosol era onnipresente e opaco. Gli unici segni visibili in esso erano una cappa scura al polo nord e un brusco mutamento di riflettività all'equatore, cosicché l'emisfero sud era nettamente più luminoso dell'emisfero nord. Titano risultò anche circondato da una sorta di strato di nebbia ad alta quota, un centinaio di chilometri circa al di sopra della parte superiore dello strato di aerosol.

La differenza di riflettività fra gli emisferi nord e sud ha ricevuto due spiegazioni alternative provvisorie. Secondo un'ipotesi sviluppata in particolare da Lawrence A. Sromovsky e Verner E. Suomi dell'Università del Wisconsin a Madison, questa differenza sarebbe una manifestazione della variazione stagionale di riscaldamento solare dello strato di aerosol causata dall'inclinazione di 26 gradi dell'asse di rotazione di Titano rispetto al piano del sistema solare. Se l'ipotesi è giusta, l'emisfero sud e l'emisfero nord dovrebbero variare periodicamente i loro valori di riflettività, con prevalenza dell'uno o dell'altro a seconda della stagione. Saturno e i suoi satelliti impiegano 30 anni a compiere una rivoluzione attorno al Sole: perciò l'alternanza dovrebbe verificarsi ogni 15 anni.

Secondo un'altra ipotesi, sviluppata da G. Wesley Lockwood, del Lowell Observatory, la differenza sarebbe il risultato di una modulazione nella produzione di particelle di aerosol dovuta alla diversa intensità delle particelle subatomiche ad alta energia del «vento» solare. Nel corso degli ultimi otto anni Lockwood ha registrato piccole variazioni nella luminosità netta di Titano. Per gran parte di tale periodo il numero delle macchie solari è andato crescendo, con conseguente intensificazione del vento solare. Continuando a osservare Titano nel corso dei prossimi anni, nella parte terminale del ciclo delle macchie solari, si potranno meglio verificare le due ipotesi: potrebbero essere esatte tutte e due.

Dopo le immagini di Titano, il *Voyager 1* fornì anche altri dati. Da un lato, lo spettrometro per l'ultravioletto a bordo della sonda rivelò a un gruppo di ricercatori diretto da Lyle Broadfoot, dell'Università della Southern California, la presenza di azoto: lo strumento scoprì nello spettro ultravioletto di Titano picchi dovuti all'emissione di radiazione ultravioletta da parte di molecole di azoto, di atomi di azoto ionizzati e di atomi di azoto non ionizzati. Lo spettro non fornì invece alcun indizio della presenza di monossido

**Questa sezione dell'atmosfera di Titano comprende due strati la cui esistenza fu scoperta da *Voyager 1*. Essi sono uno strato trasparente alla luce visibile, in cui viene assorbita la radiazione ultravioletta e, al di sotto del primo, uno strato di nebbia d'alta quota. Inferiormente alla nebbia si trova lo strato di particelle di aerosol. Si presume che, nell'intero corso della storia del sistema solare, le particelle siano andate aggregandosi in particelle più grandi, cadute via via sulla superficie. Al di sopra della superficie sono illustrate nubi di metano e piogge di metano; esse non sono confermate, ma sono probabili. La curva della temperatura in funzione della pressione *(in colore)* si fonda su un esperimento in cui l'atmosfera di Titano si interpose fra la Terra e i segnali radio trasmessi da *Voyager 1*. Secondo i dati raccolti per mezzo di quest'occultazione (unitamente a quelli raccolti dal Voyager con lo spettrometro per l'infrarosso), la temperatura alla superficie di Titano dovrebbe aggirarsi attorno a 95 kelvin e la pressione alla sua superficie sarebbe di 1500 millibar (1,5 bar). La pressione media al livello del mare sulla Terra è leggermente superiore a un bar.**

di carbonio, di argo o di neo (altre sostanze che emettono nell'ultravioletto). D'altra parte la presenza di metano e di altri idrocarburi fu suggerita quando lo spettrometro controllò l'assorbimento della luce di una stella da parte dell'atmosfera di Titano. Era in sostanza un esperimento di occultazione, con interposizione del bordo di Titano, e perciò della sua atmosfera, fra la stella e lo spettrometro.

In un secondo esperimento di occultazione la sorgente della radiazione fu la sonda stessa. In questo caso i segnali radio trasmessi verso la Terra da *Voyager 1* furono attenuati dalla rifrazione nell'atmosfera di Titano. L'attenuazione andò crescendo man mano che il Voyager spariva dietro il bordo di Titano e la densità del gas attraversato dalle onde radio aumentava. Il risultato dell'esperimento fu un profilo di densità in relazione all'altitudine nell'atmosfera di Titano. Il profilo della densità, a sua volta, fornisce un profilo di $T/\bar{\mu}$ rispetto all'altitudine, dove $T$ è la temperatura e $\bar{\mu}$ è il peso molecolare medio dell'atmosfera. L'esperimento era stato progettato da G. Leonard Tyler, della Stanford University, per vedere quale dei due modelli dell'atmosfera di Titano reggesse il confronto con l'esperienza: se quello di Danielson e Caldwell o quello di Hunten.

In definitiva l'esperimento favorì il modello di Hunten. In particolare, Von R. Eshleman della Stanford University sottolineò che il profilo di $T/\bar{\mu}$ prodotto dall'esperimento corrisponde esattamente alle previsioni del modello di Hunten per un'atmosfera ricca di azoto ad eccezione del fatto che il profilo stesso implica una pressione alla superficie inferiore rispetto al valore originario di Hunten che era di 20 bar. Il gruppo responsabile dell'esperimento di occultazione sta continuando l'analisi dei dati. Gunnar F. Lindal del Jet Propulsion Laboratory del California Institute of Technology ha comunicato i risultati più recenti del gruppo riferendo una pressione alla superficie di 1,5 bar (± 0,1 bar) e una temperatura alla superficie prossima a 94 kelvin. Lindal ha fornito anche un valore di 2575 ±2 chilometri, ricavato dall'occultazione, per il raggio di Titano.

Anche la struttura dell'atmosfera di Titano fu analizzata con spettrometri per l'infrarosso, da un gruppo diretto da Rudolf A. Hanel del Goddard Space Flight Center della NASA. La migliore concordanza reciproca fra i dati dell'occultazione e quelli dell'infrarosso indica che il peso molecolare medio dell'atmosfera di Titano è di 28,6; l'atmosfera deve comprendere perciò una quantità apprezzabile di un gas più pesante dell'azoto. (Il peso molecolare dell'azoto è 28,0.) Robert E. Samuelson del Goddard Space Flight Center ipotizza che si tratti dell'argo. Samuelson e i suoi colleghi si basano sul fatto che l'argo è relativamente abbondante nell'universo e che sarebbe gassoso alla temperatura di Titano. Inoltre l'argo (come l'azoto) è trasparente alla luce visibile e all'infrarosso vicino, cosicché non viene rilevato dalla spettroscopia nell'infrarosso.

Il fatto che lo spettrometro per l'ultravioletto di *Voyager 1* non abbia rivelato la presenza di argo significa solo che la sua abbondanza nell'atmosfera superiore di Titano dev'essere inferiore al 6 per cento. Nell'atmosfera superiore ci si attende che prevalgano i gas più leggeri; perciò il limite del 6 per cento nell'atmosfera superiore di Titano non esclude l'abbondanza netta del 12 per cento circa richiesta per dare all'atmosfera un peso molecolare medio di 28,6. Ponendo $\bar{\mu}$ uguale a 28,6 nel profilo di $T/\bar{\mu}$, si trova una temperatura alla superficie di 95±2 kelvin.

I dati trasmessi a terra da *Voyager 1* mentre si trovava in prossimità di Titano hanno rivelato quindi l'esistenza, nelle regioni esterne del sistema solare, di un satellite la cui atmosfera, come quella terrestre, è ricca di azoto. I dati ottenuti dimostrano inoltre che l'atmosfera di Titano ha una pressione alla superficie maggiore di quella della Terra. La pressione alla superficie dell'atmosfera di un corpo dipende sia dalla quantità di gas presente nella sua atmosfera sia dal grado a cui esso è compresso dalla gravitazione del corpo stesso. La gravità alla superficie di Titano è solo il 14 per cento rispetto a quella alla superficie della Terra. La pressione alla superficie di Titano (1,5 bar, secondo i dati del Voyager) significa quindi che l'atmosfera di Titano possiede una quantità di gas per area unitaria della superficie circa 10 volte maggiore di quella terrestre.

Anche lo spettrometro per l'infrarosso a bordo di *Voyager 1* registrò bande di emissione dovute a varie sostanze gassose la cui presenza nell'atmosfera di Titano non era stata accertata in precedenza. La prima sostanza identificata fu l'acido cianidrico (HCN), che potrebbe aver partecipato alle reazioni chimiche che condussero alla sintesi di composti come l'adenina sulla Terra tre miliardi di anni or sono. L'adenina è un componente del DNA ed è quindi essenziale alla vita sulla Terra. Nel corso di studi successivi Virgil G. Kunde e William Maguire del Goddard Space Flight Center e i loro colleghi confrontarono gli spettri infrarossi di Titano con spettri di laboratorio realizzati appositamente. I confronti condussero rapidamente all'identificazione di altre sei sostanze nell'atmosfera di Titano. Fra di esse vi sono idrocarburi come il propano ($C_3H_8$) e composti azotati come il cianoacetilene ($C_2HCN$). Darrell F. Strobel, del Naval Research Laboratory, ha dimostrato che queste sostanze possono derivare da reazioni implicanti azoto e metano attivate dal bombardamento dell'atmosfera di Titano da parte di fotoni ultravioletti del Sole e dagli elettroni ad alta energia intrappolati nel campo magnetico di Saturno.

## *La superficie di Titano*

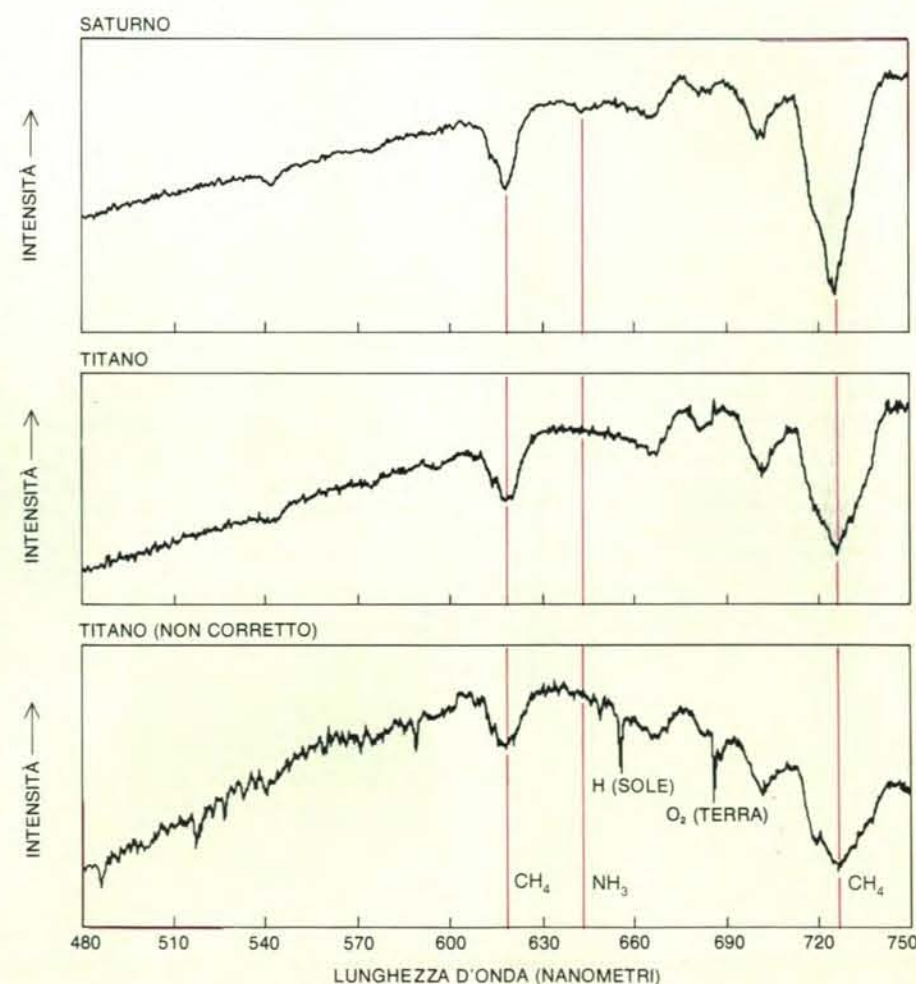
Oggi sembra chiaro che le prime speculazioni concernenti l'origine dell'aerosol di Titano erano essenzialmente corrette. I frammenti e composti molecolari prodotti dall'impatto di fotoni ultravioletti e di elettroni ad alta energia formano polimeri, ovvero catene di molecole, che vengono a trovarsi in sospensione come particelle solide nell'atmosfera. Studiando il modo in cui le particelle di aerosol riflettono la luce solare, James B. Pollack e Kathy Rages dell'Ames Research Center della NASA poterono dimostrare che la variazione nella luminosità di Titano col mutare dell'angolo fra il Sole e Titano e ciascuno dei veicoli Voyager poteva essere spiegata con l'ipotesi che le particelle di aerosol presenti nella parte alta dell'atmosfera avessero un raggio medio di 0,5 micrometri.

Si può fare l'ipotesi che queste particelle precipitino lentamente dalla sospensione e che, man mano che cadono, entrino in collisione fra loro e si aggreghino. Gli aggregati cadono più rapidamente delle particelle, cosicché l'atmosfera perde di continuo le molecole dei composti del carbonio e dell'azoto prodotti da processi che hanno luogo a grandi altitudini al di sopra della superficie di Titano. Strobel ha stimato che, da quando esiste il sistema solare, sulla superficie di Titano si dev'essere depositata una quantità di idrocarburi pari a uno strato di 0,1 - 0,5 chilometri di spessore. Contemporaneamente si dev'essere formato uno strato di composti azotati di alcune decine di metri di altezza.

Qual è la natura della superficie su cui sta cadendo questa manna dal cielo (come ama chiamarla Carl Sagan della Cornell University?) Analisi approfondite dei dati forniti da *Voyager 1* collocano la temperatura di Titano a 94±1 kelvin. Misurazioni eseguite dallo spettrometro per l'infrarosso di *Voyager 1* suggeriscono inoltre che la temperatura alla superficie vari di non più di tre kelvin fra l'equatore e i poli. La ragione di questo stato di cose risiede nell'onnipresenza dell'atmosfera densa, che assorbe gran parte della luce.

Questi valori della temperatura alla superficie consentono la presenza di metano liquido. È in effetti possibilissimo che Titano sia ricoperto da un oceano liquido, esteso sull'intero satellite, di quello che noi sulla Terra chiamiamo gas naturale.

Il metano potrebbe svolgere perciò su Titano lo stesso ruolo assolto dall'acqua sulla Terra. Alla superficie di Titano il metano è liquido. Nell'atmosfera inferiore esso è allo stato gassoso. Può darsi che l'atmosfera inferiore di Titano comprenda nubi di metano, e non è escluso che qua e là si saturi di metano, dando origine a vere e proprie piogge. A quanto pare, abbiamo trovato infine un mondo, oltre alla Terra, in cui una grande quantità di qualche composto si trova allo stato liquido sulla superficie. È un vero peccato per gli astronauti, potenziali visitatori, che la temperatura vi rimanga così vicina a 94 kelvin, ossia a -179 gradi centigradi.

Che impressione si proverebbe stando seduti su una barca in un mare di metano su Titano? La visibilità orizzontale dovrebbe essere abbastanza buona. Secondo O. Brian Toon, dell'Ames Research Center, le grandi particelle di aerosol che cadono dall'atmosfera dovrebbero essere poche e molto distanziate fra loro. La visibilità diminuirebbe, ovviamente, nel caso di un acquazzone di metano. D'altra parte la luce dovrebbe essere molto fioca. Saturno dista dal Sole quasi 10 volte più della Terra. Già questo solo fatto diminuisce di un fattore 100 l'intensità della luce solare per unità di superficie. La luce così indebolita sarebbe ancor più attenuata dallo strato di aerosol e dalle eventuali nubi di metano sovrastanti. È difficile stimare l'entità dell'attenuazione netta. Una parte dei modelli di Toon prevedono

**Questi spettri di Saturno e di Titano furono realizzati da Robert G. Danehy e dall'autore con l'ausilio del telescopio di 2,7 metri del McDonald Observatory nel Texas. Le lunghezze d'onda nel blu della luce visibile si trovano a sinistra; le lunghezze d'onda nel rosso a destra. Compaiono varie bande di assorbimento. Quelle a 619 nanometri e a circa 720 nanometri sono fra le bande di assorbimento che condussero all'identificazione del metano gassoso ($CH_4$) su Titano. Una banda meno vistosa in prossimità di 645 nanometri rivela la presenza di ammoniaca gassosa ($NH_3$) su Saturno, ma non su Titano. In generale le bande di assorbimento più intense negli spettri di Titano sono meno profonde e più larghe di quelle presenti negli spettri di Giove e di Saturno. Si ritiene che la ragione di ciò debba essere vista nella diffusione della luce ad opera delle particelle di aerosol di Titano. I due spettri presentati qui in alto sono stati divisi da uno spettro della luce riflessa dalla Luna. In questo modo le righe spettrali normalmente presenti nella luce del Sole sono state ridotte a un valore unitario e si è impedito che venissero scambiate con righe spettrali di Titano. Lo spettro di Titano in basso non è stato corretto in questo modo. Esso presenta perciò alcune righe spettrali causate dall'assorbimento di luce nell'atmosfera del Sole o nell'atmosfera della Terra.**

che la visibilità dalla barca sia press'a poco uguale a quella che si ha in una notte di luna sulla Terra, anche quando su Titano è mezzogiorno. La navigazione su Titano dovrebbe essere difficile. Il Sole e le stelle non sarebbero visibili. Inoltre una bussola non sarebbe di alcuna utilità, giacché le sonde Voyager non hanno scoperto su Titano alcun campo magnetico.

Che tipo di barca sarebbe adatto per la navigazione su Titano? Probabilmente non una barca a vela, poiché è verosimile che i venti alla superficie di Titano siano deboli, a causa della mancanza di forti differenze di temperatura da un luogo all'altro: sono tali escursioni termiche a fornire energia ai venti sulla Terra. Si potrebbe usare allora un motore fuoribordo? Qui ci imbattiamo in un curioso contrasto. Sulla Terra l'ossidante per un motore a combustione interna è presente in grande abbondanza nell'atmosfera, mentre è relativamente scarso il combustibile. Su Titano una barca galleggerebbe su un mare di carburante, mentre sarebbe scarso l'ossidante. Lo si potrebbe forse trovare scavando nel sottosuolo di Titano alla ricerca di ghiaccio d'acqua, ed estraendone poi l'ossigeno. O forse su Titano esistono continenti formati da affioramenti di ghiaccio d'acqua ricoperti da uno strato di idrocarburi e di polimeri azotati provenienti dallo strato di aerosol.

## *Come ebbe origine l'atmosfera*

Sulla scorta delle informazioni fornite dalle missioni Voyager dobbiamo tentare di capire in che modo si evolse la curiosa atmosfera di Titano. In particolare, ci troviamo di fronte a un corpo da cui l'idrogeno è sfuggito nello spazio, come avrebbe potuto dedurre Jeans 65 anni or sono. Sotto questo aspetto Titano assomiglia ai pianeti interni del sistema solare. Perché dunque Titano non ha sviluppato un'atmosfera simile a quella di Marte o di Venere, un'atmosfera ricca di anidride carbonica? La ragione risiede nella indisponibilità di ossigeno, che è intrappolato nel ghiaccio d'acqua, all'interno del corpo solido del satellite. La combinazione unica delle dimensioni di Titano e della sua temperatura ha consentito all'atmosfera di Titano di evolvere e di rimanere ancora riducente.

In generale si accetta la tesi che Titano si sia formato in una nebulosa protosaturniana, una parte isolata della nube di polvere e gas che divenne il sistema solare. Sembra quindi ragionevolmente certo anche che Titano si sia formato insieme con Saturno, i suoi anelli e i suoi altri satelliti circa 4,5 miliardi di anni or sono. La densità di Titano misurata oggi (1,9 grammi per centimetro cubo) indica che esso è formato all'incirca per il 52 per cento di roccia e per il 48 per cento di ghiacci. Le proporzioni rappresentano un leggero arricchimento in roccia rispetto alla composizione del sistema solare in generale. Il ghiaccio avrebbe avuto però una funzione cruciale per la successiva evoluzione dell'atmosera di Titano, in quanto esso deve avere intrappolato gas della nebulosa protosaturniana in misura molto maggiore di quanto non abbiano fatto le rocce.

Vent'anni or sono Stanley L. Miller dell'Università della California a San Diego previde che i satelliti ghiacciati di Saturno dovevano includere metano idratato ($CH_4 \cdot 7H_2O$), ossia metano intrappolato in ghiaccio d'acqua. La presenza oggi comprovata di metano nell'atmosfera di Titano apportò una conferma a quest'idea. La recente identificazione di vari altri gas suggerisce che anch'essi siano stati intrappolati sotto forma idratata. Per poter prevedere con sicurezza quali sostanze siano state intrappolate, si devono conoscere i valori di temperatura e di pressione per cui una data sostanza e la sua forma idratata sono in equilibrio. (Nella situazione di equilibrio il tasso con cui le molecole o gli atomi di una sostanza sfuggono dalla forma idratata è uguale al tasso con cui vengono catturati, cosicché la quantità della forma idratata non diminuisce e la forma idratata è stabile.) Si pensa che la temperatura della nebulosa protosaturniana non sia scesa di molto sotto i 60 kelvin. A quella temperatura la pressione di equilibrio per le molecole di azoto o gli atomi di argo e i loro rispettivi idrati è inferiore a $10^{-7}$ bar.

Una pressione di $10^{-7}$ bar è minore della pressione che l'azoto o l'argo fornirono probabilmente alla nebulosa protosaturniana; perciò questi gas dovrebbero essere rimasti intrappolati nei ghiacci. Invece la pressione di equilibrio per atomi di neo e per il loro idrato a 60 kelvin è di circa 40 bar. Il neo ha un'elevata abbondanza cosmica; perciò l'assenza di una quantità rilevabile di neo nell'atmosfera di Titano significa due cose: che il neo non poté essere intrappolato come idrato e, in secondo luogo, che non fu intrappolato come gas. In altri termini, la gravitazione di Titano nel periodo della sua formazione non fu abbastanza elevata da intrappolare direttamente neo dalla nebulosa protosaturniana. (Il peso atomico dell'isotopo più abbondante del neo è 20, che è quattro volte di più del limite superiore fissato dalla teoria di Jeans per la fuga da Titano alla sua massa attuale.) L'assenza di neo tende a confermare che l'atmosfera di Titano si formò dopo l'accrescimento del corpo stesso, e che l'atmosfera si formò da gas intrappolati sotto forma idratata.

**Gli spettri ripresi da *Voyager 1* nella parte infrarossa dello spettro elettromagnetico consentono d'identificare su Titano vari gas oltre al metano. In questo caso l'identificazione di cianoacetilene ($C_2HCN$) e di cianogeno ($C_2N_2$) è dimostrata da un confronto di spettri di assorbimento realizzati in laboratorio con uno spettro di emissione di Titano ripreso dal veicolo spaziale. Il confronto è valido perché le molecole di un dato gas assorbono ed emettono radiazione nello stesso insieme di lunghezze d'onda caratteristiche. Nello spettro di Titano sono indicate righe o bande spettrali di vari altri gas identificati mediante il numero d'onda, ossia il numero delle onde per centimetro. Un numero d'onda di 200 corrisponde a una lunghezza d'onda di 50 000 nanometri; uno di 1400 corrisponde a una lunghezza d'onda di circa 7100 nanometri. Gli spettri sono stati ottenuti da un gruppo diretto da Rudolf Hanel del Goddard Space Flight Center della NASA.**



**Oggi si ritiene che i gas presenti nell'atmosfera di Titano varino in abbondanza, dall'azoto molecolare (che secondo i dati del Voyager rappresenterebbe dall'82 al 94 per cento dell'atmosfera, ossia da 820 000 a 940 000 parti per milione) a tracce di idrocarburi come il metilacetilene e di sostanze azotate come il cianogeno. Potrebbero restare da scoprire vari altri componenti in tracce dell'atmosfera di Titano. Il diagramma indica un 12 per cento circa (120 000 parti per milione) del gas inerte argo. Questo 12 per cento è richiesto per far salire il peso molecolare medio dei gas componenti l'atmosfera al valore di 28,6 che risulta provvisoriamente dai dati forniti dal Voyager.**

In che modo i gas riuscirono a sfuggire dagli idrati e a raggiungere la superficie di Titano? In primo luogo la liberazione di energia potenziale gravitazionale sotto forma di calore durante l'accrescimento di Titano dovette essere sufficiente a vaporizzare una frazione dei ghiacci contenuti nel corpo. In seguito il decadimento di nuclei radioattivi all'interno del satellite dovette diventare la sua principale sorgente di calore interna. Secondo modelli proposti da Mark Lupo e da John S. Lewis del Massachusetts Institute of Technology, il riscaldamento radioattivo potrebbe essere stato sufficiente a creare una zona di acqua liquida all'interno del mantello di Titano. I gas poterono sfuggire da questo liquido.

È chiaro che vi sono modi in cui gas un tempo intrappolati nei ghiacci di Titano potevano sfuggirne e formare un'atmosfera. Indizi di tali fughe di gas dall'interno sono visibili su altri satelliti di Saturno. Le spaccature orlate da materiali più chiari del terreno circostante, su Dione, sono l'esempio più cospicuo. Dione era semplicemente troppo piccolo per conservare un'atmosfera. Altri satelliti di Saturno presentano segni di superfici giovani. Il materiale che rinnovò la superficie di tali satelliti potrebbe essere stato sospinto verso l'alto in parte dalla pressione di gas che sfuggivano da idrati.

L'azoto presente nell'atmosfera di Titano richiede un'ulteriore discussione. Negli eventi che ho descritto si suppone che l'azoto presente oggi nell'atmosfera di Titano sia entrato a far parte del satellite sotto forma di idrato nel corso del suo accrescimento. Quest'assunto richiede a sua volta che la forma dominante dell'azoto nella nebulosa protosaturniana fosse quella molecolare ($N_2$), ma la realtà potrebbe essere stata diversa. Se Ronald G. Prinn del MIT si è unito a Lewis nel suggerire che l'$N_2$ fosse la forma stabile dell'azoto nel sistema solare in formazione, Prinn e M. Bruce Fegley, Jr., anch'egli del MIT, sottolineano però anche che l'aumento di temperatura nei pressi di Giove e Saturno nel periodo della loro formazione potrebbe avere consentito la formazione di ammoniaca ($NH_3$). Se così fu e se Titano intrappolò l'azoto sotto forma di idrato anziché di $N_2$, la successiva storia di Titano dev'essere stata sostanzialmente diversa.

In particolare, calcoli eseguiti da Sushil K. Atreya e da suoi colleghi all'Università del Michigan dimostrano che agli inizi della storia di Titano si deve postulare un'epoca «calda», in cui la temperatura alla superficie dovette essere superiore a 150 kelvin. Per tutta quest'epoca dall'interno di Titano deve essere sfuggita ammoniaca, che è stata poi dissociata nell'atmosfera ad opera dei fotoni ultravioletti del Sole. In questo modo l'atmosfera deve aver perduto la sua ammoniaca e deve aver guadagnato l'$N_2$ che possiede oggi. Una temperatura di 150 kelvin non è esclusa in modo assoluto; infatti potrebbe essere stata prodotta, in linea teorica, da un effetto serra per opera dell'idrogeno e dell'ammoniaca presenti nell'antica atmosfera di Titano. L'epoca calda è nondimeno una complicazione inutile nell'ipotesi che l'azoto sia stato intrappolato da Titano nella forma di un idrato di $N_2$. Non è esclusa neppure una combinazione dei due processi, che consentirebbe di non porre limiti così precisi all'antica temperatura alla superficie.

Anche l'argo presente nell'atmosfera di Titano merita una discussione. Se esso

**Darrell F. Strobel, del Naval Research Laboratory, propone questo prospetto della chimica dell'azoto nell'atmosfera di Titano. Qui sono illustrati i principali reagenti e prodotti di reazione; sono omessi, fra gli altri, atomi e molecole di idrogeno. Fra le reazioni iniziali vi è la rottura del legame N-N nella molecola di azoto nell'atmosfera superiore. L'energia per rompere il legame è fornita dall'impatto della radiazione ultravioletta solare ($h\nu$) o di un elettrone ad alta energia ($e^*$) intrappolato in prossimità di Titano dal campo magnetico di Saturno. (Questo secondo processo fornisce all'atmosfera di Titano una quantità di energia 10 volte maggiore rispetto al primo.) Le reazioni finali comprendono la formazione delle particelle di aerosol rossastre. Le fasi finali non sono note completamente perché la composizione chimica delle particelle stesse non è ancora stabilita. Si sa però che l'acido cianidrico (HCN) è in grado di formare un polimero bruno-rossastro. Nell'illustrazione, ${}^2D$ e ${}^4S$ rappresentano stati atomici eccitati, A e $c_4'$ indicano stati molecolari eccitati ed *e* indica un elettrone liberato nell'atmosfera dalla radiazione ultravioletta solare.**

costituisce il 12 per cento dell'atmosfera, dev'essere derivato da idrati intrappolati. L'unica possibilità alternativa è che esso sia derivato dal decadimento radioattivo. Più del 99 per cento dell'argo presente nell'atmosfera della Terra è per l'appunto un prodotto del decadimento radioattivo e ha avuto origine dal decadimento dell'isotopo radioattivo potassio 40 in argo 40. Ma anche ammesso che il materiale roccioso presente su Titano avesse la stessa proporzione di potassio del materiale roccioso terrestre, il presente contenuto in argo 40 dell'atmosfera di Titano dovrebbe essere di circa 70 parti per milione, una quantità lontana da quella prevista del 12 per cento, ossia di 120 000 parti per milione. L'argo presente nell'atmosfera di Titano dev'essere in effetti argo primordiale, intrappolato sotto forma di idrato dalla nebulosa protosaturniana. Deve trattarsi perciò di argo 36, con una mescolanza del 20 per cento di argo 38.

Le pressioni di equilibrio dell'azoto e dell'argo rispetto ai loro idrati sono abbastanza simili e abbastanza basse da suggerire che le proporzioni in cui essi vengono intrappolati come idrati debbano essere press'a poco uguali alle proporzioni in cui essi poi sfuggono da quegli idrati. Un test della presenza di azoto anziché ammoniaca nella nebulosa protosaturniana comincia perciò con una stima della quantità di azoto sfuggito dal corpo solido di Titano. Questa stima deve comprendere il contenuto di azoto nell'atmosfera attuale di Titano e tener conto anche della quantità di azoto sfuggita nello spazio e della quantità che è entrata a far parte delle particelle dell'aerosol nel corso della storia del corpo celeste. Dai valori di Strobel per queste due modalità di impoverimento risulta che la quantità totale di azoto che è entrata nell'atmosfera di Titano è pari a circa 1,7 volte l'attuale contenuto di azoto dell'atmosfera, ossia al 140 per cento del contenuto dell'atmosfera di Titano.

La quantità di argo sfuggita dal corpo solido di Titano è più facile da calcolare. L'argo è inerte (non forma composti chimici) ed è troppo pesante per sfuggire nello spazio in quantità apprezzabili. La quantità di argo che è sfuggita dal corpo solido è semplicemente il 12 per cento dell'atmosfera. Il rapporto fra azoto e argo è perciò di 11,7 : 1. Il rapporto fra azoto e argo nella nube di materia che si trasformò nel sistema solare era abbastanza simile a questo, ossia di 11 : 1.

Che cosa può dirci Titano sulla Terra primitiva? Oggi si tende ad abbandonare l'opinione che la Terra avesse inizialmente un'atmosfera altamente riducente composta in prevalenza da metano, ammoniaca, idrogeno e vapore acqueo. Secondo la nuova concezione l'atmosfera terrestre doveva essere moderatamente riducente ed essere formata in prevalenza da anidride carbonica, azoto molecolare e monossido di carbonio, con non più del 10 per cento di idrogeno. Quasi tutti gli studiosi sono però d'accordo sul fatto che non doveva essere presente ossigeno libero, perché un ambiente ossidante rende estremamente difficile la formazione di molecole organiche. In un ambiente ossidante il carbonio viene rapidamente bloccato in molecole non più complesse dell'anidride carbonica.

I ricercatori hanno sottoposto a ripetute scariche elettriche in laboratorio miscugli gassosi di metano, ammoniaca, idrogeno e vapore acqueo, che dovevano simulare l'atmosfera della Terra primitiva. Essi sono riusciti a produrre, fra le altre molecole organiche, alcune di quelle che costituiscono il DNA. Quel che essi non hanno potuto produrre è qualcosa di simile al DNA, o più in generale una molecola in grado di replicare se stessa agendo come uno stampo per la formazione di un duplicato a partire dalle molecole più semplici a disposizione. Pare che il problema risieda nella mancanza, come si espresse Andrew Marvell, di «abbastanza mondo e tempo». Una provetta da laboratorio è semplicemente troppo piccola e un periodo di anni potrebbe essere troppo breve perché ci sia qualche probabilità che si formi una molecola capace di autoreplicarsi.

*Titano come laboratorio*

Sotto questo punto di vista Titano ci offre una sorta di immenso laboratorio naturale, libero da condizionamenti sperimentali e dalla possibilità di catalisi accidentale o dalla fortuita inibizione di reazioni chimiche per opera delle pareti dei recipienti di reazione o dei grassi nei sistemi di arresto di un sistema a vuoto. È inoltre un laboratorio in cui i prodotti delle reazioni che hanno avuto luogo nel corso della storia del sistema solare sono stati raccolti in particelle di aerosol che vengono conservate a basse temperature sulla superficie del satellite dove un giorno potremmo essere in grado di esaminarle a nostro piacimento. In futuro, in effetti, potremmo decidere di eseguirvi esperimenti noi stessi. Riscaldando alcuni ettari della superficie di Titano potremmo produrre una quantità di «brodo primordiale» maggiore di quella che i nostri esperimenti sulla Terra non riuscirebbero a produrre se non nell'arco di varie generazioni.

Per il momento la NASA sta pensando, per l'ultimo decennio del secolo, a una missione verso Saturno nel corso della quale una sonda dovrebbe entrare nell'atmosfera di Titano, mentre il veicolo maggiore dovrebbe restare in orbita attorno a Saturno, usando il radar per penetrare l'opacità dell'atmosfera di Titano e realizzare una mappa della superficie. Ma molto tempo prima possiamo attenderci di apprendere altre cose su Titano da un'ulteriore analisi dei dati del Voyager, dai nuovi spettrometri di grande potenza oggi disponibili sulla Terra e in particolare dagli strumenti che saranno messi in orbita attorno alla Terra con lo Space Telescope nel 1985.

Nel frattempo i veicoli spaziali Voyager continueranno le loro missioni di esplorazione. *Voyager 1* è avviato verso l'esterno del sistema solare. Se i suoi trasmettitori continueranno a funzionare, possiamo attenderci che ci riveli la posizione del confine fra il vento solare e il tenue gas del mezzo interstellare. Queste informazioni dovrebbero pervenirci entro i prossimi 10 anni. *Voyager 2* è in viaggio alla volta di Urano, che raggiungerà nel gennaio del 1986, e di Nettuno, che supererà nell'agosto del 1989. Il meccanismo di puntamento delle telecamere ha subito qualche danno nel corso del passaggio del veicolo spaziale attraverso gli anelli di Saturno. I tecnici del Jet Propulsion Laboratory confidano però che *Voyager 2* riesca a trasmettere dati importanti da Urano e da Nettuno se non vi saranno altre avarie.

I Voyager però non sono stati progettati per visite a pianeti più lontani di Saturno, cosicché i loro strumenti sono lungi dall'avere un'efficienza ottimale per informarci su corpi così lontani. Il mutamento nei piani ebbe luogo alla metà degli anni settanta quando divenne chiaro che per almeno un decennio non sarebbero state intraprese missioni specifiche dirette verso Urano e Nettuno. Dopo il successo del programma Voyager è curioso che ci si trovi a non avere altre missioni capaci di dar seguito a tale successo, tranne una che potrebbe essere annullata: il progetto Galileo, destinato a inviare una sonda nell'atmosfera di Giove per misurarne caratteristiche come composizione, pressione e temperatura. Abbiamo già perso la possibilità di inviare una missione a esaminare da vicino la cometa di Halley. In effetti stiamo regredendo all'inizio del Seicento, quando il più lontano fra i pianeti noti era Saturno e non si comprendeva che le comete attraversano il sistema solare. Ma non è il momento di una tale rinuncia. C'è ancora molto da imparare sui mondi che ci circondano prima di poter capire a fondo la storia del sistema solare e quella della Terra stessa.

# La chimica delle fiamme

*Le attuali ricerche sulla combustione sono prevalentemente rivolte ai prodotti intermedi che si formano quando gli idrocarburi bruciano producendo anidride carbonica, acqua e sostanze inquinanti in tracce*

di William C. Gardiner, Jr.

La scoperta fondamentale che permise agli scienziati di comprendere il fenomeno della combustione si ebbe nel 1774, quando Antoine Laurent Lavoisier si rese conto che l'apparente scomparsa di materia tra le fiamme non era altro che un'illusione. Lavoisier dimostrò, invece, che un componente invisibile dell'aria (che più tardi venne chiamato ossigeno) reagiva chimicamente con la materia ad alta temperatura, fornendo calore e un certo numero di prodotti di combustione. Oggi ci si basa su questa sua dimostrazione per definire un combustibile: una sostanza che reagisce con l'ossigeno dando reazione chimica esotermica (che libera calore).

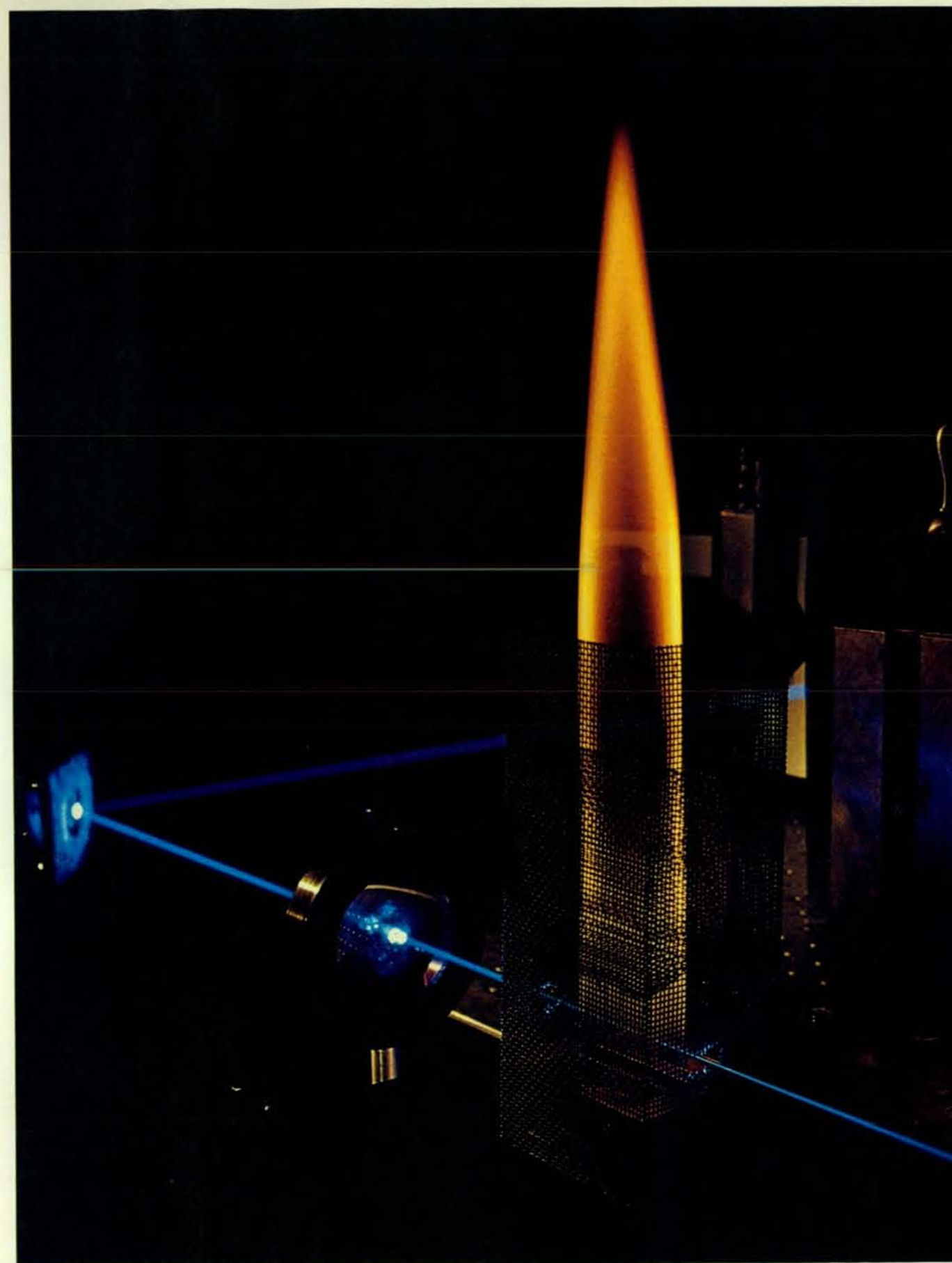
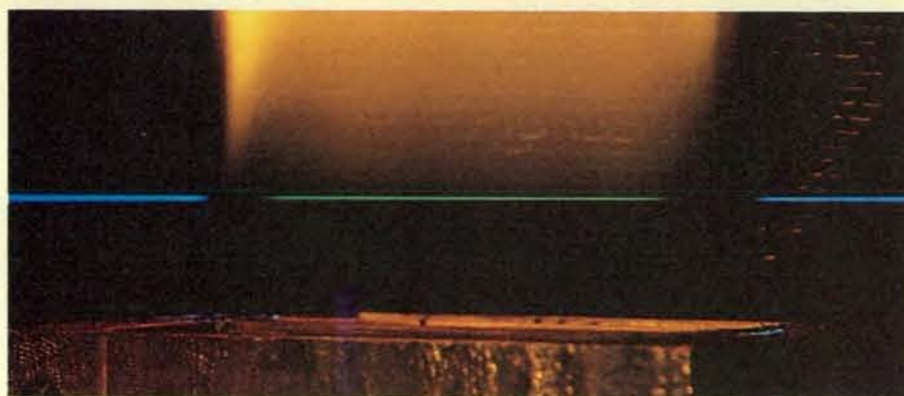
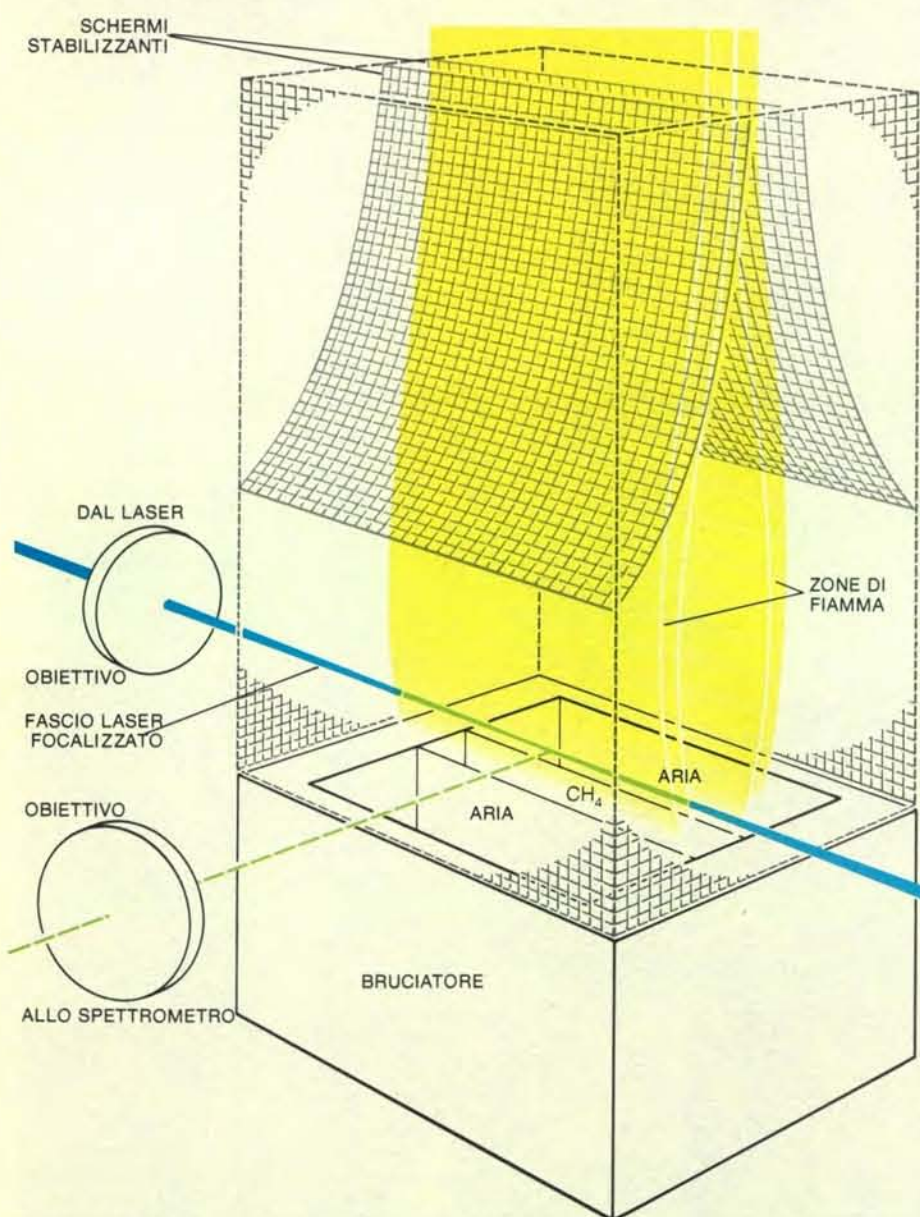
La scienza della combustione è andata assumendo sempre maggiore importanza negli anni recenti, in seguito all'aumentata consapevolezza della limitata disponibilità di combustibili fossili relativamente poco costosi e degli effetti nocivi di alcuni prodotti di combustione sull'ambiente e sulla salute dell'uomo. Avvalendosi di moderne tecniche di laboratorio è possibile rivelare non solo i prodotti finali dei processi di combustione, ma anche molte sostanze che si formano provvisoriamente nel corso della combustione. Di conseguenza si è giunti a comprendere dal punto di vista chimico il fuoco, che può essere interpretato come un intricato succedersi di eventi molecolari. L'obiettivo pratico di questo lavoro resta né più né meno quello che era per l'uomo preistorico: imparare a bruciare i combustibili più economici a disposizione nel modo più efficiente, completo e pulito possibile.

Alcuni combustibili, come la legna e lo sterco di cammello, sono troppo complessi per poterli trattare in dettaglio a livello molecolare. Altri, come l'alluminio e l'ammoniaca, richiedono più energia per la loro produzione di quante ne liberino quando bruciano e per questo motivo vengono considerati di solito poco adatti come combustibili. Gli specialisti della combustione si occupano principalmente dei combustibili che vengono bruciati comunemente come fonti energetiche: carbone, prodotti petroliferi e gas naturale. Questi combustibili non sono sostanze chimiche pure. Il carbone, che è essenzialmente costituito da carbonio elementare, presenta anche un certo numero di altri componenti, alcuni in grado di bruciare altri no. I prodotti petroliferi e il gas naturale sono miscele di idrocarburi (composti a base di idrogeno e carbonio in un rapporto approssimativo di due a uno) con tracce di altre sostanze. Oltre a questi combustibili comuni, i chimici della combustione ne hanno messo a punto altri per utilizzazioni particolari come l'acetilene per i cannelli per saldatura, metalli polverizzati per fuochi d'artificio e idrogeno per motori a razzo.

I combustibili, comunque, non bruciano spontaneamente. Una miscela di acetilene e ossigeno, per esempio, potrebbe essere immagazzinata per centinaia di anni senza che avvenga alcuna reazione percettibile. Se si espone, invece, la miscela a una fiamma o a una scintilla si ha la detonazione nel giro di un microsecondo. La spiegazione di un incompetente è semplice: si deve dapprima accendere un fuoco e questo si manterrà vivo fino a che sia il combustibile sia l'ossigeno non si siano esauriti oppure fino a quando il fuoco stesso non si sia estinto. Il compito di chi studia scientificamente la combustione è formulare in termini chimici e fisici precisi le nozioni fondamentali di ignizione, propagazione ed estinzione.

I primi chimici che si sono occupati di combustione ritenevano che il solo cambiamento di temperatura potesse spiegare i fenomeni di ignizione e propagazione. Essi pensavano che le reazioni chimiche che avvengono in un processo di combustione fossero semplicemente troppo lente per poter essere avvertite a temperature normali; se si raggiunge la temperatura di ignizione in una zona, tuttavia, la velocità alla quale la reazione chimica susseguente libera calore diventa sufficientemente alta da aumentare la temperatura nella regione adiacente fino al valore della stessa temperatura di ignizione. Questa interpretazione, che riconduce tutta la chimica della combustione a un singolo processo di liberazione di calore la cui velocità dipende dalla temperatura, non si occupa del perché di questa dipendenza dalla temperatura. Ciononostante quando il problema è formulato in termini quantitativi, l'interpretazione fornisce modelli che sono in accordo con i dati sperimentali e che risultano utili agli specialisti per progettare dispositivi per la combustione. Questi modelli spiegano in che modo la velocità di combustione dipenda dalla temperatura, ma non perché ciò avvenga, dato che non prendono in considerazione gli eventi molecolari che sono alla base del processo di combustione.

Prima di procedere ulteriormente sono necessarie alcune definizioni. Nella scienza della combustione il termine fuoco viene generalmente applicato a ogni evento accidentale o voluto di combustione nel mondo «reale» (cioè il mondo che non sia il laboratorio). In un evento di questo tipo, per esempio un fuoco in una foresta o in un forno, viene fornita aria e i prodotti di combustione e il calore vengono rimossi da flussi incostanti agenti su vasta scala. Le fiamme sono le subunità del fuoco. Esse a loro volta possono essere suddivise in due classi: fiamme premiscelate, in cui il combustibile viene miscelato con l'ossigeno prima della combustione, e fiamme a diffusione, in cui il combustibile e l'ossigeno vengono a contatto nella zona di combustione.

La benzina e l'aria (che è costituita per il 20 per cento circa da ossigeno) si miscelano nel carburatore di un motore convenzionale a combustione interna e in seguito bruciano (dopo la compressione e la ignizione innescata da scintilla nel cilindro) come fiamma premiscelata mentre la cera vaporizzata di una candela e l'aria vengono a contatto sopra a uno stoppino per bruciare come fiamma a diffusione. Le fiamme premiscelate e quelle a diffusione possono essere turbolente o laminari, a seconda delle velocità di flusso delle sostanze coinvolte. Il flusso turbolento aumenta la velocità di combustione e risulta, quindi, vantaggioso quando si desidera una combustione rapida. Il flus-

**La fiamma a diffusione laminare viene stabilizzata con l'aiuto di schermi in questa dimostrazione della tecnica della sonda laser, fotografata da Fritz Goro al Fire Research Center del National Bureau of Standards. In una fiamma di questo tipo il combustibile e l'ossidante vengono a contatto in una zona di combustione; in questo caso il combustibile è metano e l'ossidante aria. Nella fotografia in alto della pagina successiva viene mostrato un primo piano della fiamma, in cui è visibile la fluorescenza verde eccitata dal fascio laser blu che l'attraversa.**

**Questa fotografia in primo piano della fiamma, eseguita sempre da Fritz Goro, mostra la sottile linea verde di molecole di gas fluorescente eccitata dal passaggio di un fascio laser blu attraverso la zona di reazione di una fiamma a diffusione metano-aria. È stato rimosso lo schermo stabilizzante posto frontalmente. Attraverso la fiamma è vagamente visibile lo schermo posteriore.**

**La fluorescenza prodotta da una fiamma può essere analizzata per mezzo di uno spettrometro, come si vede in questo schema del dispositivo sperimentale visto nella fotografia della pagina precedente. Il metano fuoriesce da una fenditura posta al centro del bruciatore; ai lati vi sono flussi laminari di aria che esce da canali lungo le pareti del bruciatore. La combustione avviene in due zone di reazione all'interfaccia combustibile-aria. Il fascio laser è focalizzato sul lato del combustibile della zona di reazione vicino al bruciatore. Gli esperimenti sulla formazione di fuliggine al Fire Research Center sono stati compiuti da W. Gary Mallard e Kermit C. Smyth.**

so laminare, o uniforme, è più semplice da descrivere matematicamente ed è di solito più adatto per esperimenti di laboratorio. In tutti i tipi di fiamma sopra ricordati il combustibile è già presente sotto forma di vapore, prima che abbia inizio la combustione. Se non lo è, allora ci si trova di fronte a una situazione fisica più complicata: la combustione eterogenea, come quella che ha luogo sulla superficie di un pezzo di carbone che sta bruciando, o un processo combinato di vaporizzazione e combustione, come avviene in un motore diesel o in un motore di un aereo a reazione.

Tra tutti questi tipi di fiamma solo quella premiscelata laminare offre la possibilità di effettuare uno studio quantitativo dei processi chimici nonostante i processi fisici concomitanti, in particolare il moto di diffusione di calore e di materia. Tuttavia, anche la zona di reazione relativamente semplice di una fiamma di un becco Bunsen non è ottimale per lo studio della chimica della combustione; il modello di flusso attorno a questa fiamma è bidimensionale, cioè ha una dimensione in più rispetto alle semplici equazioni di flusso. Inoltre, la zona di reazione è molto circoscritta e, quindi risulta difficile eseguire analisi chimiche all'interno della fiamma. Per studiare la struttura di fiamme in laboratorio si può generare un modello di flusso monodimensionale sostituendo il tubo di un becco Bunsen con un contenitore di fiamma grosso e piatto dotato di un tappo poroso. La zona di reazione può allora essere estesa riducendo la pressione della miscela combustibile-ossidante a un valore inferiore alla pressione atmosferica.

Nelle fiamme premiscelate a bassa pressione è possibile seguire l'andamento delle reazioni chimiche con appropriate sonde che rivelano la composizione e la temperatura del gas nella zona di combustione. Sonde intrusive, come i tubi di quarzo per il prelievo di gas e le termocoppie, forniscono informazioni a risoluzione spaziale sulla composizione e la temperatura anche se alterano il flusso, spesso in modo significativo e difficile da descrivere matematicamente. Sonde ottiche, come i fasci laser, hanno il vantaggio di non disturbare la fiamma stessa, ma lo svantaggio di far deviare il percorso del fascio luminoso a causa dei gradienti di densità presenti nella fiamma. Sul finire degli anni cinquanta sono state sperimentate numerose fiamme utilizzando un certo numero di metodi resisi disponibili quando venne sviluppata la tecnologia in grado di generare fiamme stabili a bassa pressione.

La variazione di concentrazione dei componenti molecolari di una fiamma può essere rappresentata, in funzione della distanza dal bruciatore, da un insieme di curve conosciute come profili di fiamma. L'analisi matematica di questi profili, compiuta per la prima volta da Joseph O. Hirschfelder e Charles F. Curtiss dell'Università del Wisconsin a Madison negli anni cinquanta, permette di dividere i cambiamenti di concentrazione molecolare in due classi: cambiamenti dovuti alla diffusione e cambiamenti dovuti alle reazioni chimiche. Si può, allora, ricavare la velocità netta di reazione chimica per ogni sostanza presente nella fiamma; l'insieme risultante di profili di velocità di reazione chimica viene utilizzato come descrizione chimica di base della fiamma. Ancora prima che i chimici della combustione sviluppassero le tecniche per ottenere profili di reazione attendibili, tuttavia, essi sapevano che nessun insieme di profili di fiamma, anche se completo, sarebbe stato sufficiente per identificare gli eventi molecolari responsabili dei profili. Essi erano già consapevoli che la varietà degli eventi molecolari era molto più grande della quantità dei profili di fiamma che avrebbero mai potuto sperare di raccogliere.



| FIAMMA | REAZIONE CHIMICA | TEMPERATURA (KELVIN) | ENERGIA EMESSA (JOULE PER GRAMMO) |
|---|---|---|---|
| IDROGENO-OSSIGENO | $2H_2 + O_2 \rightarrow 2H_2O$ | 3100 | 24 000 |
| METANO-OSSIGENO | $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$ | 3000 | 10 000 |
| METANO-ARIA | $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$ | 2200 | 2700 |
| OTTANO-OSSIGENO | $2C_8H_{18} + 25O_2 \rightarrow 16CO_2 + 18H_2O$ | 3100 | 9900 |
| ACETILENE-OSSIGENO | $2C_2H_2 + 5O_2 \rightarrow 4CO_2 + 2H_2O$ | 3300 | 11 800 |
| CIANOGENO-OSSIGENO | $C_2N_2 + O_2 \rightarrow 2CO + N_2$ | 4800 | 6300 |
| GASSOGENO-ARIA | $2CO + 4H_2 + 3O_2 \rightarrow 2CO_2 + 4H_2O$ | 2400 | 4100 |
| METILIDRAZINA-TETROSSIDO D'AZOTO | $CH_6N_2 + N_2O_4 \rightarrow 2H_2O + CO_2 + 2N_2$ | 3000 | 7500 |

**Le condizioni della fiamma dipendono sia dal combustibile sia dall'ossidante. Le temperature elencate in questa tabella per le reazioni chimiche si riferiscono a fiamme che bruciano a pressione atmosferica. A pressioni superiori dovrebbe essere soppressa la dissociazione dei prodotti di combustione, con conseguenti temperature di fiamma superiori. I valori forniti per l'energia liberata in ogni reazione si riferiscono alla quantità di energia emessa per grammo complessivo di combustibile e ossidante (compresi i componenti inerti nel caso dell'aria), dopo che i gas prodotti sono stati raffreddati a temperatura ambiente. Le fiamme di idrogeno e di cianogeno sono importanti in laboratorio per studi di spettroscopia ad alta temperatura. Il gassogeno è una miscela combustibile relativamente economica preparata da carbone e acqua. La metilidrazina viene bruciata con tetrossido d'azoto come ossidante nei motori che controllano l'assetto della navetta spaziale; diversamente dalla maggior parte delle combinazioni combustibile-ossidante, queste due sostanze danno luogo al contatto a una ignizione istantanea.**

Le molecole della maggior parte dei combustibili hanno troppi atomi che partecipano alla combustione per poter pensare che il processo sia un evento che procede di concerto. Immaginate la confusione che si avrebbe se gli otto atomi di carbonio e i 18 atomi di idrogeno di una molecola di ottano ($C_8H_{18}$) dovessero staccarsi uno dall'altro e ricombinarsi tutti in una volta con le molecole circostanti di ossigeno biatomico ($O_2$) formando i nuovi legami chimici dell'anidride carbonica ($CO_2$) e dell'acqua ($H_2O$). Nessun combustibile brucia in questo modo. Al contrario, la scissione di molecole di combustibile e la formazione di prodotti di combustione procedono attraverso lunghe sequenze di passaggi; ogni passaggio comprende solo una piccola ridisposizione di legami chimici.

Un passaggio di questo tipo è noto come reazione elementare; le varie molecole che si formano in questo stadio vengono chiamate collettivamente prodotti intermedi di reazione. L'insieme di tutte le reazioni elementari che, prese globalmente, sono responsabili della trasformazione chimica finale costituiscono il meccanismo di reazione. L'equazione che descrive la reazione chimica globale che ha luogo in una fiamma non fornisce indizi sui singoli cambiamenti molecolari; le equazioni che descrivono le reazioni elementari, tuttavia, rappresentano eventi chimici reali a livello molecolare. Solamente quando si conoscono tutte le reazioni elementari importanti si può descrivere, con una certa accuratezza, la trasformazione del combustibile in prodotti di combustione, in termine di ridistribuzioni di atomi nelle molecole.

La chiave per la comprensione delle varie fasi della combustione è la scoperta delle reazioni elementari. Per acquisire questa conoscenza non è sufficiente, tuttavia, identificare reazioni elementari particolari come eventi chimici possibili, ma si deve anche sapere qual è la probabilità che le molecole coinvolte reagiscano effettivamente quando i loro percorsi si incrociano, dato che la maggior parte dei loro incontri non dà luogo a una reazione chimica. Questa probabilità può invertirsi in funzione della velocità relativa con la quale le molecole si urtano. Di solito quanto più violenta è la collisione tanto più probabile è il verificarsi di una reazione chimica. Dal momento che sia la frequenza, sia la velocità degli incontri molecolari dipendono dalla temperatura del gas, la probabilità di reazione aumenta, spesso molto bruscamente, con l'aumento della temperatura.

Lo studio delle probabilità di reazione e la loro dipendenza dalla temperatura rientrano nell'area della cinetica chimica, un campo in cui è consuetudine esprimere queste probabilità in termini di costanti di proporzionalità, dipendenti dalla temperatura, chiamate coefficienti di velocità di reazione. La velocità di reazione è data dal prodotto tra il coefficiente di velocità e la concentrazione di ognuno dei reagenti. Dato che i valori dipendenti dalla temperatura dei coefficienti di velocità della maggior parte delle reazioni della fiamma non sono stati misurati, nell'ambito delle temperature delle fiamme si fa ricorso di solito a estrapolazioni. Fortunatamente la teoria dei coefficienti di velocità è stata sperimentata in modo esauriente ed è (con alcune riserve) abbastanza affidabile da permettere estrapolazioni accurate.

Gli obiettivi della chimica della combustione sono pertanto quelli di identificare le reazioni elementari di una fiamma per determinare il coefficiente di velocità di ogni reazione in funzione della temperatura e di riunire queste informazioni in modelli di fiamma che consentano previsioni controllabili. Per arrivare a queste previsioni si utilizzano programmi al calcolatore in cui le equazioni che descrivono le velocità di diffusione e di reazione di cambiamenti di concentrazione di tutte le molecole presenti nella fiamma vengono combinate con le equazioni che descrivono il flusso dei gas reagenti.

Nel ricercare le reazioni elementari delle fiamme, i chimici della combustione sanno in anticipo i tipi di reazione che dovranno identificare. Agli inizi di questo secolo si era scoperto che le sostanze intermedie che partecipano alla maggior parte delle reazioni chimiche sono presenti in quantità estremamente basse, che di solito non sono rivelabili. Questa è una caratteristica dei prodotti intermedi che sono particolarmente reattivi rispetto alle altre molecole presenti e che partecipano facilmente a reazioni elementari che, a loro volta, producono altri prodotti intermedi. Un esempio molto importante nella chimica della fiamma è la reazione elementare che converte il monossido di carbonio in anidride carbonica: $CO + OH \rightarrow CO_2 + H$. L'intermedio di reazione che entra nella reazione elementare, il

radicale ossidrile (OH), fornisce un altro intermedio di reazione, l'idrogeno atomico (H), con il risultato che non vi è alcuna variazione nel numero di intermedi. (Il termine radicale viene usato dai chimici per designare un frammento reattivo delle molecole.)

L'atomo di idrogeno partecipa in seguito ad altre reazioni elementari, alcune delle quali possono a loro volta dare un radicale ossidrile, rendendo possibile la ripetizione della reazione con monossido di carbonio. A causa della ripetizione degli stessi eventi molecolari, questo tipo di meccanismo di reazione viene chiamato reazione a catena; le reazioni elementari sono chiamate stadi di propagazione della catena e gli atomi e i radicali che partecipano alla propagazione della catena sono noti come centri di catena. Un numero molto basso di centri di catena può portare a reazioni chimiche.

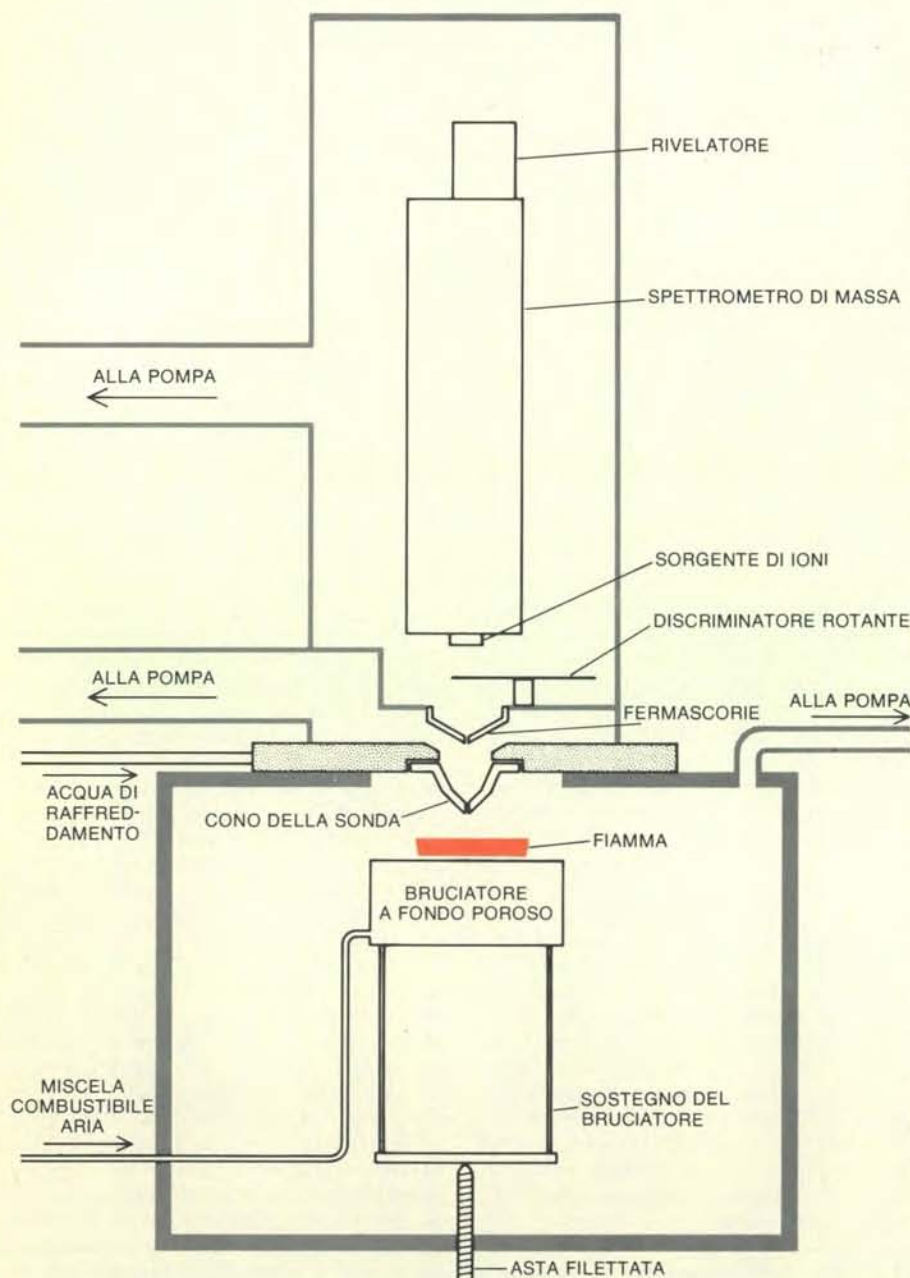
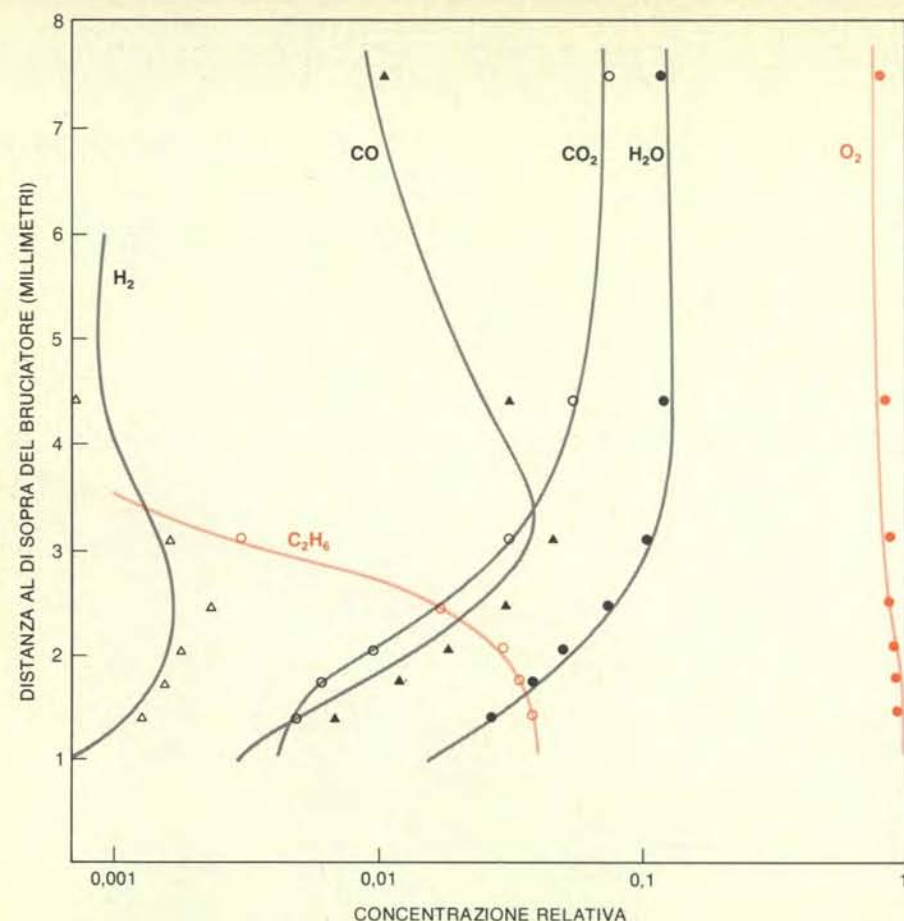
Gli stadi di propagazione della catena non sono i soli responsabili delle fiamme. È stato scoperto negli anni trenta che sono indispensabili anche gli stadi di inizio, termine e ramificazione della catena. Il premio Nobel 1956 per la chimica fu assegnato congiuntamente a Cyril N. Hinshelwood dell'Università di Oxford e a Nikolai N. Semenov dell'Istituto di fisica chimica di Mosca per il loro contributo alla dimostrazione di come le principali caratteristiche delle reazioni di ossidazione potessero essere spiegate in termini di meccanismi di reazione a catena ramificata. (Nel periodo in cui si svolgevano queste ricerche i chimici disponevano solamente delle previsioni su quelle che potevano essere le reali reazioni elementari, mentre mancavano quasi del tutto informazioni sui valori dei coefficienti di velocità.)

Uno stadio di inizio della catena è una reazione elementare in cui uno o due molecole normalmente stabili reagiscono per formare uno o due centri di catena. Un esempio è fornito dalla reazione $C_3H_8 \rightarrow C_2H_5 + CH_3$ dove la molecola stabile di propano ($C_3H_8$) si decompone spontaneamente, quando viene portata a un livello energetico elevato da collisioni con altre molecole in un gas caldo, dando luogo alla formazione di due radicali $C_2H_5$ e $CH_3$, ognuno dei quali può fungere a sua volta da centro di catena.

Uno stadio di termine della catena è una reazione elementare che presenta l'effetto opposto. Per esempio, nella reazione $H + OH + N_2 \rightarrow H_2O + N_2$ due centri di catena, un atomo di idrogeno e un radicale ossidrile, incontrano simultaneamente la molecola d'azoto chimicamente stabile ($N_2$) e formano un'altra molecola chimicamente stabile, l'acqua. In questo caso la molecola d'azoto agisce come un accompagnatore, assorbendo parte dell'energia che viene liberata quando si forma il nuovo legame chimico della molecola d'acqua.

La caratteristica chimica distintiva delle fiamme, tuttavia, è la presenza di stadi di ramificazione della catena: reazioni elementari in cui il numero di centri di catena aumenta da uno a due o probabilmente anche a tre. La reazione di ramificazione della catena più importante è $H + O_2 \rightarrow O + OH$, dove l'idrogeno atomico reagisce con l'ossigeno molecolare dando ossigeno atomico e un radicale ossidrile.

Questa teoria di estrema importanza è conosciuta da circa mezzo secolo. Per sviluppare una descrizione dettagliata della chimica della fiamma, tuttavia, erano necessari due fattori ulteriori. Il primo era rappresentato da un insieme di procedimenti matematici per trovare gli effetti combinati di tutte le reazioni elementari sulle proprietà chimiche e fisiche osservabili. Esprimere i processi fisici di diffusione, conduzione termica, reazione chimica e flusso in forma matematica era solo il punto di partenza; le equazioni differenziali simultanee risultanti presentano due difficoltà, una ovvia e l'altra più sottile.

La prima difficoltà consiste nel fatto che una descrizione esatta di diffusione e di conduzione termica richiede che si sia in possesso di alcuni parametri (chiamati coefficienti di trasporto a più componenti) che non possono essere misurati in laboratorio. Loro descrizioni approssimate possono essere fatte solamente facendo

**Per seguire l'andamento delle reazioni di combustione in fiamme che bruciano a bassa pressione viene usata una sonda campione. L'espansione rapida del gas che entra nella sonda a forma di cono ha un effetto raffreddante che «congela» le reazioni della fiamma. Il campione di gas catturato dalla sonda può essere analizzato esternamente alla zona di reazione con metodi analitici convenzionali, come la spettrometria di massa impiegata in questo dispositivo progettato da Joan C. Biordi, Charles P. Lazzara e John Papp presso lo US Bureau of Mines Research Center di Pittsburgh. Sono, invece, richiesti metodi ottici per ottenere profili di fiamme di sostanze molto reattive.**

**I profili di fiamma calcolati *(curve)* possono essere messi a confronto con misurazioni di laboratorio *(punti di riferimento)* per fiamme laminari a bassa pressione. I profili mostrati qui sono stati ottenuti da Jürgen Warnatz della Technische Universität di Darmstadt per un meccanismo di reazione comprendente 58 reazioni elementari, tenendo conto della diffusione di tutte e 20 le sostanze coinvolte nella reazione globale. I dati sperimentali, per una fiamma di etano-ossigeno che brucia a una pressione pari a un decimo di quella atmosferica, sono stati ottenuti da Robert M. Fristrom, William H. Avery e C. Grunfelder della Johns Hopkins University. Le analisi al laser di fiamme analoghe, condotte da James H. Bechtel e dai suoi collaboratori al General Motors Research Laboratory, hanno confermato i profili ottenuti con la spettrometria di massa.**

affidamento su ipotesi di accuratezza discutibile. Per fortuna una recente ricerca di modelli al calcolatore ha dimostrato che le approssimazioni fatte nella descrizione della diffusione e della conduzione del calore hanno solo effetti secondari sui risultati.

La seconda difficoltà consiste nel fatto che le equazioni differenziali così ottenute hanno una proprietà chiamata «rigidezza»; questa fa sì che alcune variabili della concentrazione vengano mantenute a valori quasi costanti per mezzo di reazioni molto rapide, mentre altre cambiano. In questo modo, se l'insieme delle equazioni differenziali simultanee deve essere risolto da un calcolatore digitale con metodi convenzionali sono necessari tempi estremamente lunghi. Negli anni settanta sono state messe a punto speciali tecniche al calcolatore per risolvere equazioni differenziali complesse. Questo risultato è frutto per la maggior parte della ricerca pionieristica di Charles W. Gear dell'Università dell'Illinois a Urbana. L'avvento di tecniche numeriche rigidamente stabili ha permesso di ottenere modelli di fiamme monodimensionali, che tenessero debitamente conto della chimica della fiamma, su grossi calcolatori digitali.

Una simulazione al calcolatore fornisce buoni risultati solamente in funzione dei dati che gli vengono forniti, in questo caso i coefficienti di velocità per tutte le reazioni elementari. Dato che le fiamme stesse sono troppo complesse per definire questi coefficienti, essi vanno cercati in ambienti meno complessi. Di conseguenza la seconda necessità fu quella di misurare sperimentalmente i coefficienti di velocità delle reazioni elementari. Come?

Nonostante l'importanza della combustione in fase gassosa a scopo energetico, gli scienziati che studiano le reazioni dei gas impersonificano tuttora la «piccola scienza»: ristretti gruppi (talvolta una sola persona) con piccoli laboratori e scarse dotazioni di fondi, dispersi nel mondo, che si incontrano personalmente soprattutto ai margini di convegni scientifici, il cui scopo principale è quello di avere alcune informazioni sulla cinetica di base dei gas. Quando venne sviluppata per la prima volta la teoria delle reazioni della fiamma, i cinetisti studiavano le reazioni di ossidazione in bulbi di vetro, dove si poteva far procedere lentamente le reazioni mantenendo temperature e pressioni inferiori rispetto a quanto avviene normalmente nelle fiamme. In questo modo il risultato di lunghi anni di scrupoloso lavoro ha permesso di identificare un certo numero di reazioni elementari e di dedurre per queste i coefficienti di velocità. Tutti questi sforzi, comunque, hanno trascurato le reazioni a catena centrali, poiché sono troppo veloci per essere studiate con questi metodi.

Lo sviluppo di tecniche di laboratorio sofisticate ha permesso di studiare anche le reazioni elementari più veloci. Nessuna tecnica sperimentale singola e nessun laboratorio, in particolare, hanno dominato il campo; le nuove conoscenze sulle reazioni elementari di combustione sono il frutto del lavoro di molti ricercatori che hanno sfruttato tecniche diverse. Per rendersi conto di questa ricerca descriverò tre esperimenti basati su tre tecniche che hanno fornito i contributi più importanti. L'insieme delle tre reazioni elementari studiate in questi esperimenti costituisce la reazione principale a catena della fiamma di idrogeno-ossigeno.

Un'onda d'urto che attraversa un gas lo colpisce ripetutamente. La temperatura può variare di migliaia di kelvin in un miliardesimo di secondo. In un tubo d'urto, un dispositivo di laboratorio in cui le onde d'urto di intensità controllata vengono fatte passare attraverso un gas campione, le reazioni chimiche, che hanno luogo in seguito al riscaldamento da urto, possono essere controllate per mezzo di metodi analitici rapidi. Nel 1966 David Gutman e Garry L. Schott del Los Alamos Scientific Laboratory hanno utilizzato un tubo d'urto per generare miscele di idrogeno e ossigeno, scaldate dall'urto, sufficientemente diluite con argo, un gas chimicamente inerte, per mantenere una temperatura costante durante la reazione. La crescita dei centri di catena poteva essere seguita per mezzo della spettroscopia a fluorescenza in una scala dei tempi dell'ordine dei microsecondi.

Gutman e Schott si munirono di una sonda analitica aggiungendo, al gas campione, monossido di carbonio, che emette una luminescenza blu con un'intensità che è proporzionale alla concentrazione degli atomi di ossigeno presenti. La velocità alla quale aumentava l'intensità della luminescenza veniva estrapolata fino al valore zero del rapporto $O_2/H_2$ dove la velocità di ramificazione della catena dovrebbe dipendere esclusivamente dalla velocità della reazione $H + O_2 \rightarrow OH + O$. Il risultato fu la determinazione, nell'ambito di temperature comprese tra 1100 e 1700 kelvin, del coefficiente di velocità della reazione elementare responsabile della ramificazione di catena nelle fiamme di idrogeno-ossigeno.

Atomi reattivi possono essere ottenuti a basse temperature decomponendo molecole per mezzo di scariche elettriche.

È possibile studiare reazioni veloci di atomi prodotti in questo modo miscelandoli con altri gas e facendo fluire la miscela rapidamente oltre una stazione di osservazione situata a valle. Nel 1968 Arthur A. Westenberg e Newman deHaas del Laboratorio di fisica applicata della Johns Hopkins University ottennero una piccola quantità di ossigeno atomico in una scarica a microonda di ossigeno molecolare diluito con il gas inerte elio. La miscela di ossigeno atomico ed elio venne aggiunta a un flusso di idrogeno che proveniva da un tubo mantenuto a una temperatura di reazione desiderata. La perdita di atomi di ossigeno, a valle del punto di miscelazione, dovuta alla reazione con le molecole di idrogeno veniva misurata con uno spettrometro a risonanza di spin elettronico. Si poté così determinare, in un ambito di temperature comprese tra 500 e 900 kelvin, il coefficiente di velocità della seconda reazione elementare delle fiamme di idrogeno-ossigeno, e cioè $O + H_2 \rightarrow OH + H$.

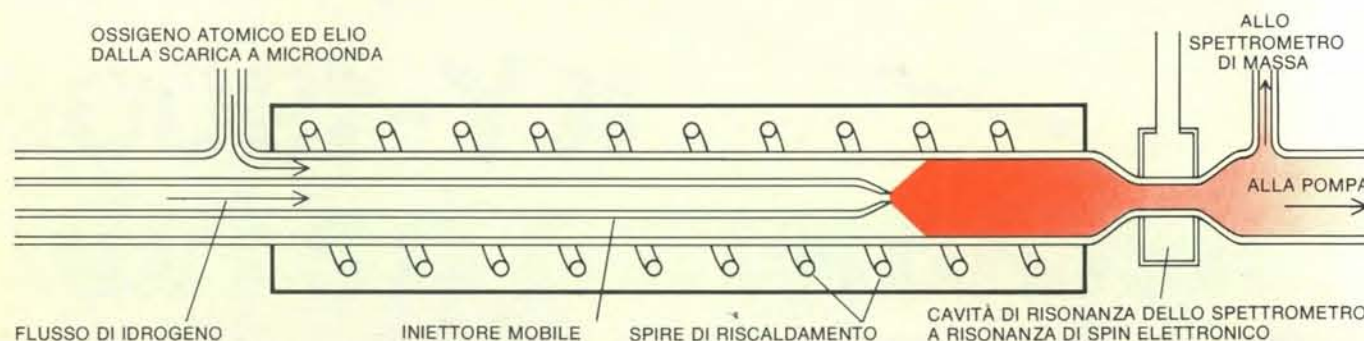
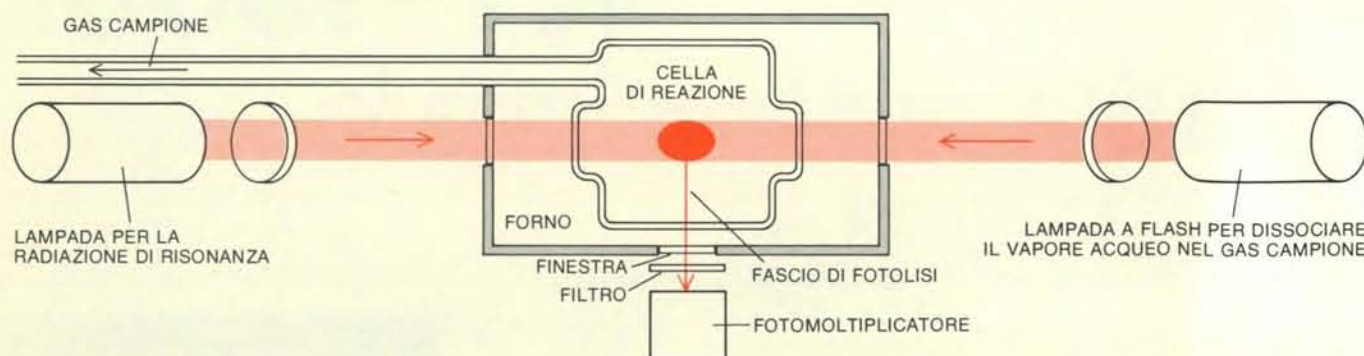
La radiazione ultravioletta fornisce un altro metodo per decomporre molecole stabili. Nel 1980 Frank P. Tully e Akkihebbal R. Ravishankara del Georgia Institute of Technology hanno messo a punto un'apparecchiatura nella quale una piccola quantità di vapore acqueo disciolto in idrogeno a una temperatura di reazione desiderata poteva essere irradiata con un intenso lampo di radiazione ultravioletta. I radicali ossidrile formatisi in seguito al lampo potevano quindi reagire con le molecole di idrogeno nella terza reazione elementare della fiamma di idrogeno-ossigeno, e più precisamente $OH + H_2 \rightarrow H_2O + H$. L'abbassamento di concentrazione di OH veniva seguito per mezzo della spettroscopia analitica ultravioletta. Si determinava in questo modo il coefficiente di velocità della reazione in un ambito di temperatura compreso tra 500 e 1000 kelvin.

Dozzine di esperimenti come questo hanno procurato ai chimici un insieme di coefficienti di velocità della fiamma di idrogeno-ossigeno, rendendo pertanto possibile un modello accurato di queste fiamme al calcolatore. Oltre alle tre reazioni elementari discusse in precedenza, per costruire modelli che descrivano anche i particolari di minore importanza della reazione idrogeno-ossigeno se ne possono includere circa altre 20, la maggior parte delle quali riguardano il ruolo dei prodotti intermedi secondari perossido d'idrogeno ($H_2O_2$), radicale perossidrile ($HO_2$) e ozono ($O_3$).

Dal punto di vista scientifico le fiamme di idrogeno-ossigeno rivestono interesse come prototipi dei processi di combustione, ma la loro sola applicazione al di fuori del laboratorio è nei motori a razzo. Le vere fiamme bruciano idrocarburi. Che cosa si sa di esse?

Una descrizione chimica comune di una fiamma di idrocarburo la farebbe assomigliare a un libro che presenta una introduzione quasi completa (le reazioni delle molecole di combustibile), un capitolo conclusivo completo (l'ossidazione di $H_2$ e CO) e, nel mezzo, solo frammenti. Il numero di capitoli del libro dovrebbe aumentare in funzione del numero di atomi di carbonio nelle molecole di combustibile, tranne che per i combustibili più semplici, quelli con uno o due atomi di carbonio, che condividono vie di reazione comuni e, quindi, presentano lo stesso meccanismo di combustione. I capitoli iniziali potrebbero derivare dai combustibili e dagli intermedi conosciuti: $CH_4$, $CH_3$, $CH_2$, CH, $C_2H_6$, $C_2H_5$, $C_2H_4$, $C_2H_3$, $C_2H_2$, $C_2H$, $CH_3O$, $CH_2O$, CHO, $C_2H_4O$, $C_2H_3O$, $C_2H_2O$...

Modelli sperimentali di fiamme di idrocarburi si ottengono combinando le reazioni elementari, utilizzando sia i coefficienti di velocità misurati sia quelli presunti. Esse possono essere esaminate in un contesto vario di eventi sperimentali: velocità di propagazione di fiamma, misurazioni di ignizione in un tubo d'urto, profili di concentrazione molecolare in un reattore a flusso e così via. Anche per gli idrocarburi a catena più corta, tuttavia, la lista di reazioni elementari considerate importanti per una ragione o per l'altra è composta da più di 100 ingressi, e il solo modo per conservare tracce del processo è quello di servirsi di diagrammi di flusso.

**Esperimenti col tubo d'urto sono stati compiuti riscaldando un gas campione con un'onda d'urto che scende in un lungo tubo e osservando successivamente la reazione conseguente nel gas campione caldo. Il dispositivo mostrato in questa illustrazione è stato utilizzato da David Gutman e Garry L. Schott del Los Alamos Scientific Laboratory nel 1966 per determinare il coefficiente di velocità della reazione $H + O_2 \rightarrow OH + O$. Il gas campione era una miscela di molecole biatomiche di idrogeno ($H_2$), ossigeno ($O_2$) e monossido di carbonio (CO) fortemente diluita nel gas inerte argo. L'onda d'urto era generata dallo scoppio di un sottile diaframma metallico che separava il gas guida ad alta pressione (in questo caso $H_2$) dal gas campione a bassa pressione. La reazione chimica cominciava quando l'onda d'urto veniva riflessa dalla parete terminale del tubo, portando il gas campione, in un miliardesimo di secondo, alla temperatura di reazione. L'intensità della luminescenza prodotta dalla reazione CO + O, che è proporzionale alla concentrazione degli atomi di ossigeno nel gas campione, veniva registrata da un tubo fotomoltiplicatore attraverso una finestra di osservazione posta nella parete terminale del tubo d'urto. La velocità della reazione chimica, misurata su una scala di microsecondi, veniva dedotta dalla velocità con la quale aumentava l'intensità della luminescenza.**

**Esperimenti di flusso a scarica permettono di studiare reazioni rapide facendo scorrere il gas reagente ad alta velocità al di là di un punto d'osservazione. In questo esperimento, eseguito nel 1968 da Arthur A. Westenberg e Newman deHaas della Johns Hopkins, l'ossigeno atomico era generato in una scarica a microonde e miscelato con idrogeno molecolare fatto fluire in un tubo riscaldato. Il coefficiente di velocità della reazione $O + H_2 \rightarrow OH + H$ era determinato misurando le concentrazioni dei reagenti e dei prodotti con uno spettrometro a risonanza di spin elettronico situato a valle del punto di miscelazione. Le misurazioni venivano anche compiute con uno spettrometro di massa.**

**Esperimenti di fotolisi vengono eseguiti iniziando la reazione con un intenso lampo di radiazione ultravioletta. Il dispositivo qui illustrato è stato utilizzato da Frank P. Tully e Akkihebbal R. Ravishankara del Georgia Institute of Technology in uno studio del 1980 sulla reazione $OH + H_2 \rightarrow H_2O + H$. Gli ossidrili (OH) sono stati ottenuti per fotodissociazione di una traccia di vapore acqueo all'interno della cella di reazione riscaldata per mezzo di una lampada a flash *(a destra)*. La velocità di reazione è stata determinata controllando l'intensità della fluorescenza proveniente da una piccola frazione di ossidrili eccitata a un livello energetico radiativo dalla lampada a flash di sinistra.**

**Le reazioni elementari di una fiamma di idrogeno avvengono in presenza delle sostanze intermedie H, O e OH. Questa sequenza di reazioni mostra un tipico effetto di «ramificazione di catena»: per ogni atomo di idrogeno che entra nella sequenza vengono prodotti tre atomi di idrogeno. Questo fenomeno è quello che fornisce alle reazioni di combustione la loro velocità esplosiva. Sono stati misurati i coefficienti di velocità di tutte queste reazioni. A basse temperature e alte pressioni vi sono altre reazioni che producono e consumano $H_2O_2$, $HO_2$ e ozono.**

**Gli idrocarburi bruciano attraverso una complessa successione di fasi di combustione che coinvolge un numero così alto di reazioni elementari che è necessario un diagramma di flusso per descrivere il corso globale del meccanismo di reazione. Questo diagramma di flusso relativamente semplice mostra le trasformazioni principali che hanno luogo nella combustione di idrocarburi a catena corta come metano ($CH_4$), etano ($C_2H_6$) ed etilene ($C_2H_4$). Gli intermedi O, H, $H_2$ e OH e i prodotti di combustione $CO_2$ e $H_2O$ sono omessi dal diagramma. Non vengono neppure mostrate le ulteriori reazioni dell'acetilene ($C_2H_2$) e le reazioni degli idrocarburi intermedi $C_2H$, $C_2H_3$ e $CH_2$, alcune delle quali portano alla formazione di idrocarburi più complessi e di fuliggine.**

I particolari dei modelli di fiamma di idrocarburi sono ancora in discussione, ma molte caratteristiche generali sono chiare. In primo luogo, la maggior parte delle reazioni elementari sono di un tipo già conosciuto da tempo nella cinetica chimica. Le reazioni di trasferimento di atomi come $OH + C_2H_6 \rightarrow H_2O + C_2H_5$ costituiscono il principale attacco sui combustibili a base di idrocarburi. Molecole di combustibile più grosse, tuttavia, solitamente non raggiungono intatte la zona di reazione, poiché vengono decomposte dal calore più in fretta di quanto diffondano. Il tipo più comune di decomposizione comporta la perdita contemporanea di due atomi di carbonio, come nella decomposizione dell'ottano tramite la reazione $C_8H_{18} \rightarrow C_2H_5 + C_6H_{13}$. La maggior parte dei composti intermedi con un numero dispari di atomi di idrogeno, come il radicale $C_2H_5$, perdono rapidamente uno di essi per formare intermedi della classe di idrocarburi chiamati olefine, in questo caso etilene ($C_2H_4$). Altri possono perdere atomi di idrogeno per formare intermedi della classe degli etini; un rappresentante tipico di questo gruppo è l'acetilene ($C_2H_2$).

Si è scoperto che olefine ed etini subiscono nelle fiamme alcuni tipi di reazioni elementari che non sono note nella cinetica chimica tradizionale. Per esempio, Jürgen Warnatz dell'Università di Darmstadt, basandosi su modelli della velocità di fiamme di acetilene, è giunto di recente alla conclusione che i radicali $CH_2$ che si formano quando gli atomi di ossigeno reagiscono con l'acetilene attraverso la reazione $O + C_2H_2 \rightarrow CH_2 + CO$ possono «andare in pezzi» reagendo con ossigeno molecolare secondo la reazione $CH_2 + O_2 \rightarrow CO_2 + H + H$. Queste trasformazioni complesse di legami chimici in singole reazioni elementari avvengono raramente al di fuori delle fiamme.

I modelli al calcolatore della combustione di metano ($CH_4$) comprendono già più di 100 reazioni elementari. C'è dunque qualche speranza di costruire un modello per l'ottano ($C_8H_{18}$), ipotizzando anche di avere a disposizione calcolatori con notevole capacità di calcolo? La risposta è che non c'è alcuna speranza se il modello deve tenere precisamente conto di tutti i prodotti intermedi: il numero di reazioni elementari richieste aumenta in proporzione geometrica con le dimensioni della molecola di combustibile e diventa eccessivo molto prima che venga raggiunto l'ottano. Se, tuttavia, il modello di combustione dell'ottano servisse solo a valutare quantitativamente gli eventi chimici centrali a livello molecolare, i calcolatori oggi disponibili potrebbero svolgere questo lavoro raggruppando i prodotti intermedi a seconda della loro struttura e le reazioni a seconda delle loro caratteristiche. Anche se si possono determinare i coefficienti di velocità per un modello di questo tipo, la successiva generazione di chimici della combustione dovrà sempre scoprirli in laboratorio.

La combustione riversa inquinanti nell'atmosfera: ossidi di azoto e zolfo, idrocarburi non completamente bruciati e particelle sotto forma di fuliggine. I tecnici ritengono che sia possibile eliminare le emissioni inquinanti regolando le condizioni di combustione, selezionando combustibili a basso tenore di azoto e zolfo, ricorrendo al pretrattamento dei combustibili o al trattamento finale dei gas di combustione. Tuttavia la soluzione scientificamente più valida sarebbe quella di scoprire quali reazioni elementari producano o eliminino gli inquinanti.

L'aspetto meglio conosciuto dell'inquinamento da combustione è la formazione di ossidi di azoto aventi formula generica $NO_x$ in fiamme molto calde (a temperature superiori a 2000 kelvin). Il monossido d'azoto (NO) è apparentemente il solo ossido d'azoto che si forma direttamente nelle fiamme; il biossido d'azoto ($NO_2$) si forma successivamente attraverso reazioni più lente con l'ossigeno atmosferico, che hanno luogo anche a temperatura ambiente. In fiamme molto calde, l'azoto molecolare presente nell'aria perde le sue caratteristiche di gas inerte e viene scomposto dagli atomi di ossigeno nella reazione $O + N_2 \rightarrow NO + O$. Gli atomi di azoto attaccano rapidamente l'ossigeno molecolare formano monossido d'azoto e ossigeno atomico nella reazione $N + O_2 \rightarrow NO + O$, completando in questo modo una catena di due reazioni conosciuta come meccanismo di Zeldovich (da Yakov B. Zeldovich dell'Istituto di fisica chimica di Mosca, che per primo lo propose nel 1947). Misurazioni dirette nel tubo d'urto del coefficiente di velocità per la reazione $O + N_2 \rightarrow NO + N$ riportate da Jamie Monat, Ronald K. Hanson e Charles H. Kruger, Jr., della Stanford University nel 1980, fanno ritenere che sia possibile riprodurre, attraverso questo procedimento, un modello accurato della formazione di monossido d'azoto.

A temperature sufficientemente basse da inattivare le reazioni di Zeldovich, altre reazioni possono produrre monossido d'azoto. Charles Fenimore del General Electric Research Laboratory ha dimostrato che tracce di NO compaiono molto presto nelle fiamme e l'hanno portato a coniare il termine «NO immediato». Non è chiaro allora quale prodotto intermedio sia sufficientemente reattivo e abbastanza abbondante per attaccare $N_2$. Un'ipotesi è che il radicale CH sia in grado di farlo, forse attraverso la sequenza di reazioni: $O + C_2H_2 \rightarrow CH + CHO$; $CH + N_2 \rightarrow HCN + N$ e infine $N + O_2 \rightarrow NO + O$. Si pensa che anche l'acido cianidrico (HCN) sia un importante composto intermedio quando il monossido d'azoto si forma da azoto combustibile (il termine generico per un certo numero di composti contenenti azoto che si trovano nel petrolio in piccole concentrazioni). Le reazioni che portano dall'acido cianidrico al monossido d'azoto non sono ancora note. Tuttavia, la maggior parte del monossido d'azoto che si forma all'inizio nelle fiamme non sfugge mai nell'atmosfera, ma viene convertito in azoto molecolare. Come? Nessuno lo sa. C'è ovviamente ampio spazio per ulteriori ricerche sulla formazione e la rimozione nelle fiamme di composti di formula $NO_x$.

Le conoscenze di base di reazioni che coinvolgono il monossido d'azoto hanno condotto a un metodo ingegnoso per la rimozione del monossido d'azoto dal gas di postcombustione. Il metodo che è stato sviluppato su scala commerciale da Richard K. Lyon, della Exxon Research and Engineering Corporation, richiede l'iniezione di ammoniaca ($NH_3$) nel gas di postfiamma a un punto particolare del processo di raffreddamento. Questa porta rapidamente alla formazione di radicali $NH_2$, soprattutto attraverso la reazione $NH_3 + OH \rightarrow H_2O + NH_2$. Una delle reazioni elementari dell'$NH_2$, riportata nel 1972 da Manfred Gehring, Karlheinz Hoyermann, Helmuth Schacke e Jürgen Wolfrum dell'Università di Gottinga, è $NH_2 + NO \rightarrow N_2 + H_2O$. Nello stadio successivo entrambi gli atomi di azoto delle molecole reagenti si trasformano in azoto molecolare, eliminando completamente il monossido d'azoto.

Un composto a base di zolfo che entra in una fiamma forma rapidamente la molecola stabile di anidride solforosa ($SO_2$), dato che l'SO, l'analogo solforato dell'NO, reagisce direttamente con l'ossigeno attraverso la reazione $SO + O_2 \rightarrow SO_2 + O$, anche a temperatura ambiente. I chimici che studiano le reazioni dello zolfo che danno luogo a SO e $SO_2$ hanno il problema che lo zolfo in fase gassosa genera una stupefacente schiera di composti reattivi e stabili, tutti acidi e tutti inquinanti atmosferici indesiderabili. C'è, quindi, scarso desiderio pratico di pervenire all'identificazione di reazioni elementari specifiche dello zolfo nelle fiamme. Il controllo dell'inquinamento da zolfo significa rimuovere l'elemento in tutte le sue forme prima della combustione o successivamente rimuovere tutti le anidridi dello zolfo.

Per finire c'è la fuliggine, la cui formazione e ossidazione sono correntemente i due processi più attivamente studiati nella chimica della combustione. Una volta il principale uso del petrolio si aveva nelle lampade a cherosene, che richiedono la formazione intermedia di fuliggine per fornire luminescenza e la conseguente ossidazione della fuliggine per evitare l'annerimento del camino della lampada. Al giorno d'oggi la luce può essere prodotta più convenientemente con l'elettricità, di modo che la fuliggine nella maggior parte delle fiamme è solo un disturbo. I fattori che aumentano la produzione di fuliggine sono ben noti: alto rapporto combustibile-aria, composti contenenti combustibile con basso rapporto idrogeno-carbonio, ignizione irregolare e scarsa miscelazione di combustibile e aria. Sembra probabile che la maggior parte dei combustibili liquidi proverrà da scisti bituminosi e da carbone liquefatto che sono per loro natura combustibili a basso tenore d'idrogeno. Potranno essere bruciati su larga scala senza seppellirci tutti nella fuliggine?

La fuliggine si forma in tre stadi. In tutte le fiamme di idrocarburo c'è una diminuzione nel rapporto idrogeno-carbonio delle molecole di idrocarburo e dei radicali man mano che vengono decomposti dal calore e dall'attacco chimico. Le ulti-

**La composizione molecolare di una tipica fiamma di laboratorio che produce fuliggine varia in funzione della distanza dal bruciatore. Questo insieme di profili logaritmici di concentrazione, per una fiamma di acetilene-ossigeno, riunisce i risultati di misurazioni condotte da Ulrich Bonne, Klaus H. Homann e Heinz-Georg Wagner dell'Università di Gottinga, da Douglas B. Olson e Hartwell F. Calcote degli Aerochem Research Laboratories e da Jack B. Howard e dai suoi collaboratori al MIT. Il gruppo di composti intermedi etichettato come «precursori di fuliggine» consiste di una varietà di idrocarburi che presentano da 20 a 50 atomi di carbonio. A differenza delle molecole di idrocarburi poliaromatici, come $C_{14}H_8$, i cui profili di fiamma continuano a salire, i precursori di fuliggine scompaiono non appena compare la fuliggine.**

**La struttura allungata della fuliggine è evidente in questa microfotografia a trasmissione elettronica eseguita da Robin Stevenson dei General Motors Research Laboratories. Particelle di fuliggine di questo tipo si formano in una fiamma di idrocarburo ricca di combustibile dall'agglomerazione di un aerosol di minute sfere di carbonio, contenenti quantità variabili di idrogeno e di idrocarburi intrappolati. In fiamme di colore giallo, ma che non emettono fumo nero, l'aerosol di fuliggine viene ossidato dai gas di fiamma più velocemente di quanto non si agglomeri. Questa particella è stata raccolta su un substrato di carbonio amorfo dal tubo di scappamento di un motore diesel. L'immagine, ingrandita 110 000 volte, è stata sottoesposta per aumentarne il contrasto.**

# La visione dei colori nei pesci

*Le acque in cui vivono hanno colori che vanno dal blu all'infrarosso vicino. I pigmenti della retina acquisiti dai pesci per adattarsi a questi ambienti permettono di studiare l'evoluzione dell'occhio*

di Joseph S. Levine ed Edward F. MacNichol, Jr.

Le cellule fotosensibili della retina sono i prolungamenti esterni del sistema visivo di un animale. Le cellule catturano fotoni di luce visibile, li contano, facendo così scattare una complessa serie di eventi chimici e nervosi che possono dar luogo alla sensazione visiva. In molti animali la funzione di raccolta dati delle cellule si è sviluppata per sfruttare le distinzioni possibili sulla base di colore e luminosità, poiché in molti ambienti la visione cromatica è un mezzo efficiente per ottenere informazioni essenziali per la sopravvivenza. Tuttavia, nelle acque profonde il numero di fotoni che raggiungono un organismo è molto ridotto, e il colore della luce ambientale è limitato a una stretta banda dello spettro. Il sistema di fotorecettori presente nella retina di un organismo che vive nelle acque profonde si è di conseguenza sviluppato più per aumentare la sensibilità alla luce disponibile, che per fornire la visione specifica dei colori. Tra il buio delle acque profonde e la luce intensa della superficie esiste una gamma continua di ambienti luminosi che variano sia per intensità sia per colore. In un tale laboratorio naturale ci si imbatte in una enorme varietà di sistemi visivi.

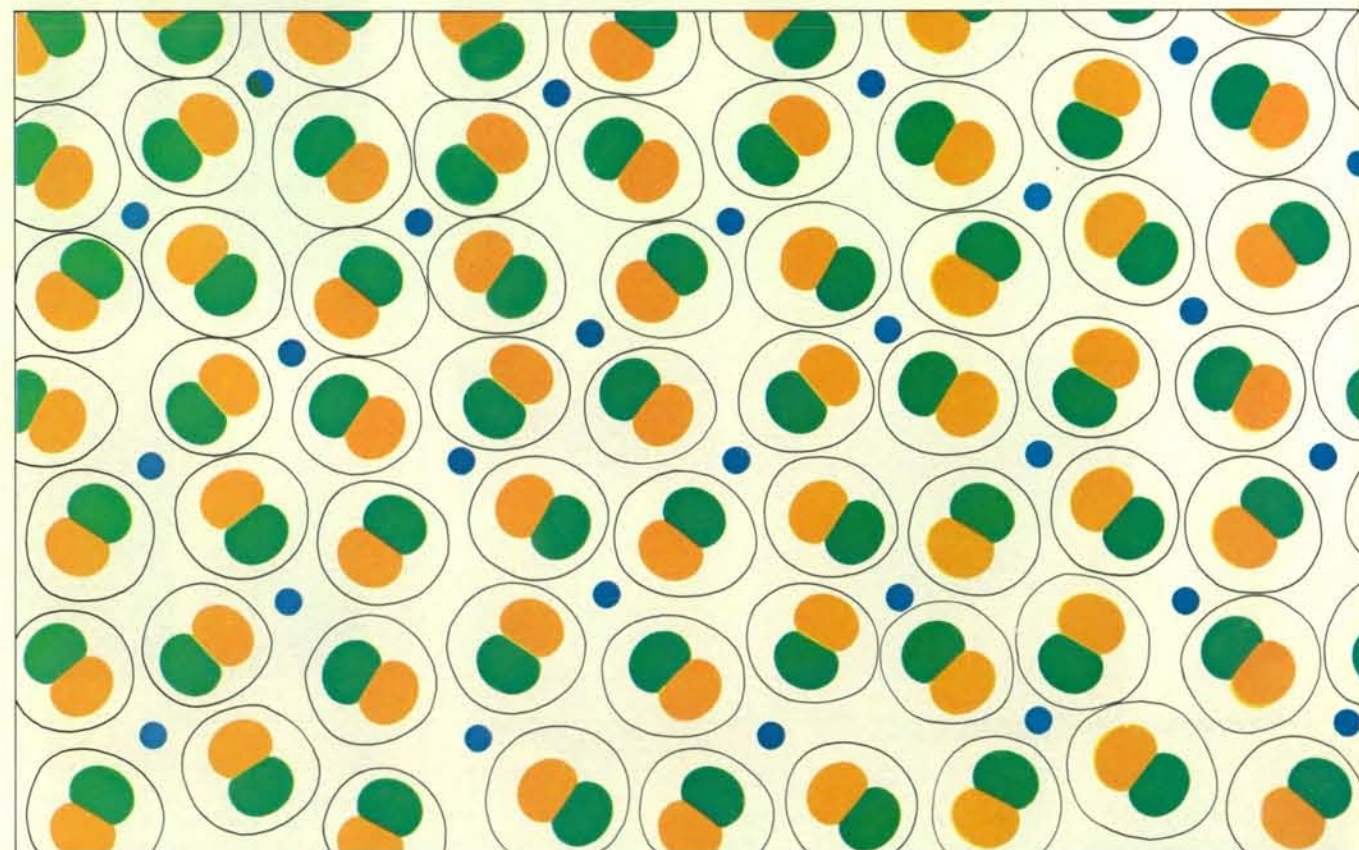
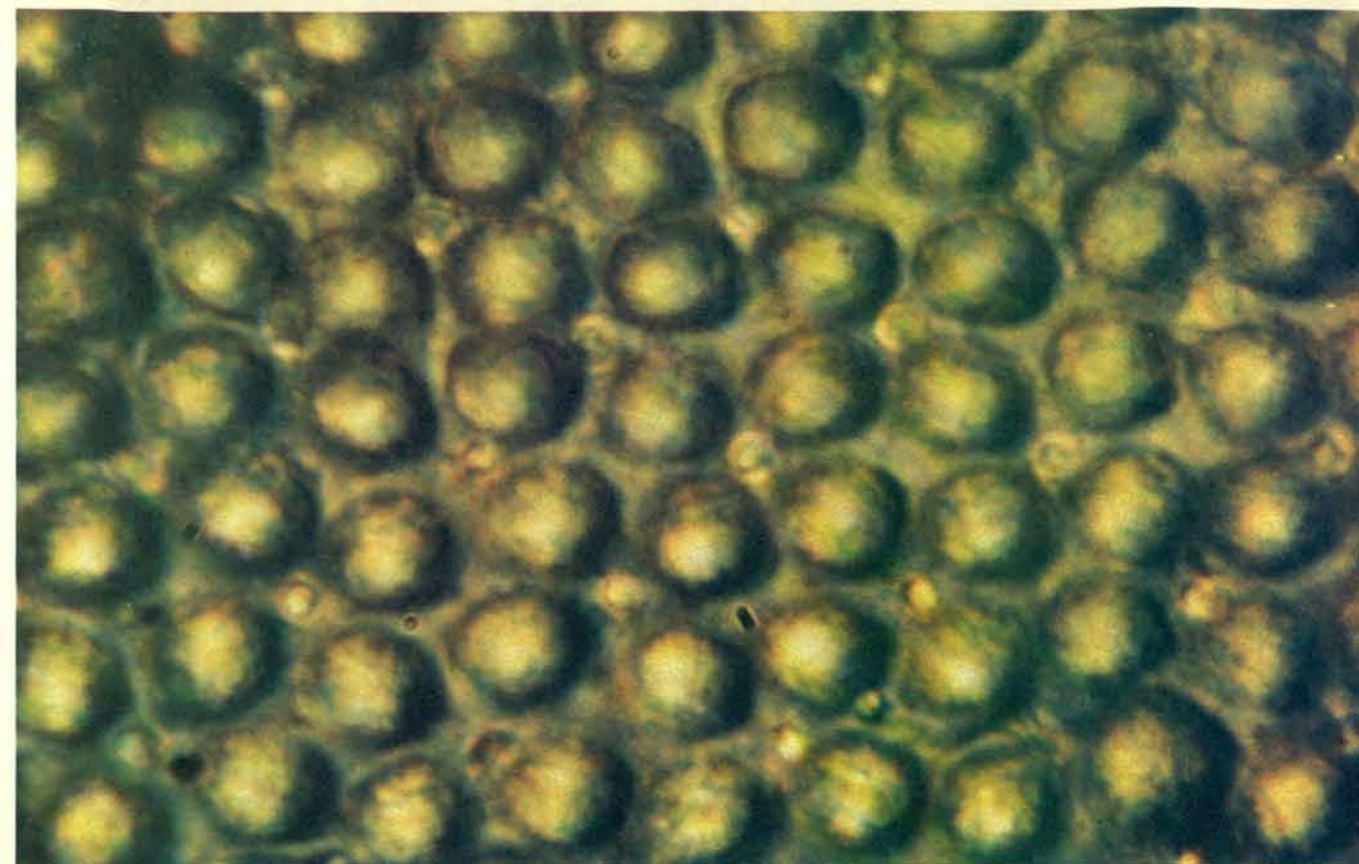
Parecchi anni fa cominciammo uno studio approfondito sui fotorecettori dei teleostei (pesci dotati di uno scheletro osseo), nella speranza di capire in che modo i sistemi visivi complessi potrebbero essersi evoluti dai loro precursori. La ricostruzione evolutiva del sistema visivo non può basarsi sui reperti fossili poiché i tessuti molli dell'occhio e le vie nervose non si conservano. Perciò scegliemmo di studiare rappresentanti tipici provenienti da una varietà di famiglie di teleostei per potere esaminare alcune delle relazioni tra i sistemi visivi, il comportamento, evidentemente da essi influenzato, e l'ambiente. Se si conoscono le relazioni evolutive tra le specie viventi di teleostei in diversi campi, un esame dei loro sistemi visivi e della loro ecologia visiva potrebbe condurre alla ricostruzione dei possibili stadi evolutivi della visione.

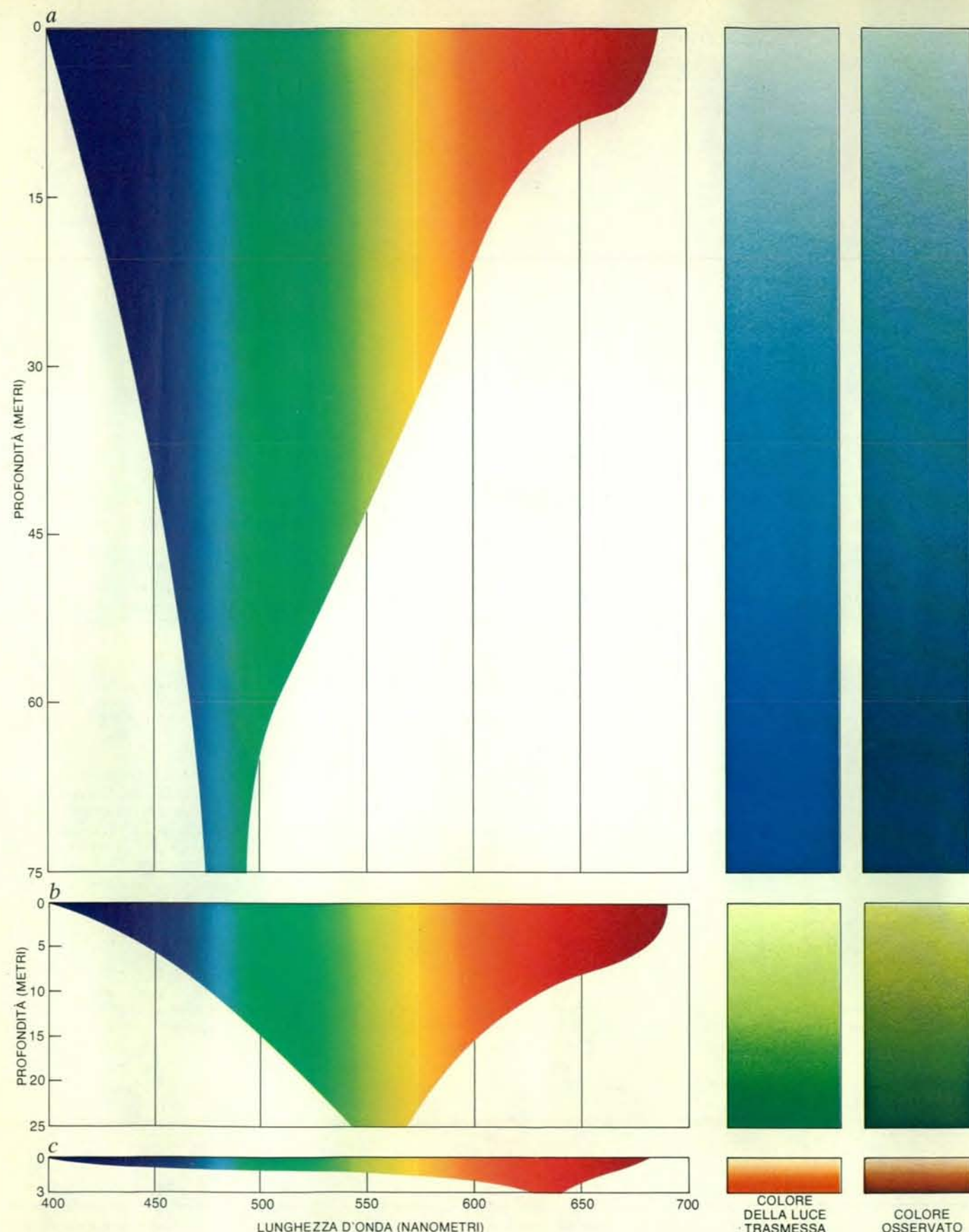
Si potrebbe supporre che questo tipo di ricostruzione sia relativamente semplice. Si potrebbe ipotizzare, per esempio, che pesci di cui si sa, grazie ai reperti fossili, che hanno subito piccoli cambiamenti evolutivi in tempi recenti, conservino ancora sistemi visivi relativamente primitivi. Quindi si potrebbe anche assumere che specie evolutesi in tempi relativamente recenti possano incorporare sistemi di visione più elaborati di quelli trovati nelle specie più antiche. Entrambe le ipotesi, tuttavia, non tengono conto della possibilità che gli occhi possano essersi evoluti indipendentemente da altri organi, poiché gli occhi stessi possono portare vantaggi selettivi per talune attività vitali come la ricerca del cibo, la fuga dai predatori, la scelta di un compagno. I vantaggi selettivi di un sistema di visione piuttosto che di un altro, in un particolare ambiente per un particolare repertorio di comportamenti, possono indurre cambiamenti negli occhi di un pesce che per il resto rimane inalterato dalle forze evolutive. I teleostei illustrano tali possibilità evolutive con particolare chiarezza.

Un'esplosiva ondata evolutiva ha interessato da 17 000 a 30 000 specie di teleostei che occupano una vasta gamma di habitat acquatici. Anche singole famiglie di teleostei, aventi diversi tipi di comportamenti dipendenti dalla vista, occupano habitat che abbracciano quasi l'intera gamma di zone luminose nel mare e nei laghi d'acqua dolce. I loro occhi possono differire così tanto per dimensioni, forma, organizzazione della retina e sensibilità ai colori, che non si sospetterebbe mai, prendendo in considerazione solo gli occhi, che tali specie siano strettamente correlate. D'altra parte, specie indipendenti che popolano la stessa zona luminosa possono presentare notevoli somiglianze nei loro sistemi visivi. Per questo motivo, i sistemi visivi primitivi, ancora presenti in teleostei viventi, non possono essere distinti chiaramente da sistemi degenerati o specializzati.

Nonostante questa difficoltà, lo studio ecologico comparato può dare una risposta ad alcune importanti domande di carattere evolutivo. Quale combinazione di vantaggi selettivi e di vincoli attivi favorisce il possesso di sistemi visivi capaci di scegliere in una vasta gamma di colori e di intensità luminose? In quali circostanze il possesso di un sistema visivo così complesso è un ostacolo più che un vantaggio? Sebbene si sappia abbastanza poco circa i vincoli operanti sulla visione, che sono correlati specificamente alla nutrizione, alla fuga o all'accoppiamento, sono molte le conoscenze attualmente disponibili sulle condizioni luminose che prevalgono nell'ambiente sottomarino.

La radiazione solare che raggiunge la superficie dell'acqua è costituita da fotoni di tutte le lunghezze d'onda del visibile e da fotoni delle regioni dell'infrarosso e dell'ultravioletto dello spettro elettromagnetico. Le energie dei fotoni di luce visibile corrispondono a colori o lunghezze d'onda: dal viola, con lunghezza d'onda di 400 nanometri, al rosso scuro intenso con lunghezza d'onda di 700 nanometri. Quando la luce passa attraverso l'acqua la sua intensità diminuisce e la perdita di intensità varia con il colore. Le masse d'acqua dolce e salata che non contengono molta materia organica assorbono la luce violetta e rossa molto più di quanto assorbano la luce di lunghezze d'onda intermedie. Quindi, gli oceani tropicali e i laghi puliti come il Crater Lake nell'Oregon e il Lake Malawi nell'Africa orientale diventano sempre più blu con la profondità, poiché vengono assorbiti tutti i colori tranne il blu. In queste acque la luce blu può penetrare fino a profondità di oltre 75 metri, mentre la luce rossa e la luce viola sono praticamente assenti a profondità superiori ai 25 metri circa.

A differenza di questi habitat di «acque chiare», la maggior parte dei laghi della zona temperata, i corsi d'acqua e le acque costiere contengono sia fitoplancton giallo-verde sia materia organica disciolta derivante dalla putrefazione di piante e animali. In tali acque viene assorbita luce di tutte le lunghezze d'onda più fortemente che nelle acque limpide,



**I fotorecettori nella retina di *Gymnocorymbus ternetzii* formano un regolare disegno a mosaico. Nella microfotografia in alto, eseguita dagli autori, i coni, ingranditi 1000 volte, deputati alla visione dei colori, appaiono come si vedrebbero dal retro della retina del pesce. (La maggior parte dei bastoncelli, che sono più sensibili dei coni alla luce, sono stati rimossi dal campione per chiarezza.) Nella retina di *G. ternetzii* sono presenti tre tipi di coni: coni singoli contenenti il pigmento visivo sensibile alla luce blu e coni doppi, costituiti da cellule separate, contenenti il pigmento sensibile sia alla luce verde sia a quella arancione. Nel disegno in basso il colore della luce maggiormente assorbita da ogni pigmento è associato alla posizione dei pigmenti nella retina. Il colore delle cellule nella microfotografia è un artefatto introdotto dalla tecnica della microscopia.**

**La trasmissione della luce da parte dell'acqua dipende dal colore o lunghezza d'onda della luce. Nelle acque chiare degli oceani e dei laghi *(a)* la luce diventa sempre più monocromatica e blu man mano che aumenta il suo tragitto. Nelle acque dolci, che contengono materia organica verde *(b)*, la luce di qualsiasi lunghezza d'onda viene assorbita più rapidamente che non nelle acque chiare, ma la luce diventa più verde man mano che aumenta la lunghezza percorsa. Nei fiumi, nelle paludi e negli stagni in cui sono presenti grosse quantità di prodotti di decomposizione vegetali e animali *(c)*, l'assorbimento è rapido e la distribuzione spettrale della luce si sposta verso il rosso. Tali acque vengono chiamate nere poiché l'occhio umano è abbastanza insensibile alla luce di elevata lunghezza d'onda; una definizione meno antropomorfica potrebbe essere acque infrarosse. Le profondità alle quali si ha la massima penetrazione della luce sono tipiche, ma variano largamente.**

ma le intensità relative di ogni colore, alle varie profondità, sono abbastanza differenti da quelle che si hanno nelle acque limpide. La luce viola a lunghezza d'onda inferiore e la luce blu sono quelle assorbite più fortemente; man mano che aumenta la profondità, sempre più domina la luce verde o giallo-verde, con lunghezza d'onda di circa 540-560 nanometri. All'occhio umano l'acqua si presenta verde.

Gli acquitrini, le paludi e le «acque nere» dei fiumi contengono gli stessi composti che assorbono come i laghi e i torrenti le piccole lunghezze d'onda, insieme a tannini, lignine e altri prodotti derivanti da una più completa decomposizione dei vegetali. Gli effetti combinati della materia in sospensione e della materia disciolta possono portare all'assorbimento di tutte le lunghezze d'onda della luce, entro tre metri dalla superficie: la lunghezza d'onda in gran parte trasmessa è di 600 nanometri o più, cioè nel rosso. All'occhio tale acqua appare normalmente scura o rosso-bruna.

La luce che attraversa l'acqua viene inoltre diffusa in tutte le direzioni da particelle sospese o dalle molecole stesse di acqua. In seguito alla diffusione la luce deve propagarsi ulteriormente nell'acqua per raggiungere un determinato punto. Inoltre, a causa della diffusione ogni linea di vista (visuale) nell'acqua è una fonte di fotoni, cosicché l'acqua stessa sembra colorata uniformemente. Questa colorazione è chiamata «luminosità di sfondo» e gli effetti congiunti della diffusione e dell'assorbimento fanno sì che essa vari con la direzione della linea di vista.

La luminosità di sfondo è un fenomeno ben noto ai subacquei, ma solo verso la metà degli anni settanta ne vennero eseguite misurazioni spettroradiometriche in habitat marini tropicali. William N. McFarland della Cornell University e Frederick W. Munz dell'Università dell'Oregon confermarono quantitativamente le prime osservazioni secondo cui la luce lungo le linee di vista orizzontali è molto più monocromatica di una luce lungo una linea di vista verticale, vicino alla superficie. McFarland e Munz riconobbero nella loro scoperta la causa principale dell'evoluzione dei sistemi visivi acquatici. Diversamente dagli animali terrestri, che devono distinguere il cibo, i predatori e i compagni da qualunque cosa possa accadere nell'ambiente durante l'osservazione, i pesci di solito devono individuare gli oggetti su una luminosità di sfondo, che mantiene un colore relativamente costante lungo una particolare linea di vista.

**La diffusione della luce da parte delle molecole d'acqua e delle particelle sospese fa sì che il colore dominante della luce trasmessa vari con la direzione della linea di vista (visuale). Poiché ogni punto su ogni linea di vista è una sorgente di fotoni, l'acqua stessa appare colorata in modo diverso in tutte le direzioni. La colorazione determina uno sfondo di luce ambiente contro il quale vanno visti tutti gli oggetti, e interpone una luminosità velata tra l'osservatore e l'oggetto. L'illustrazione mostra come la colorazione cambi con la direzione in quattro microhabitat nelle «acque nere» di un fiume: in prossimità della superficie e sul fondo, in acque poco profonde e profonde.**

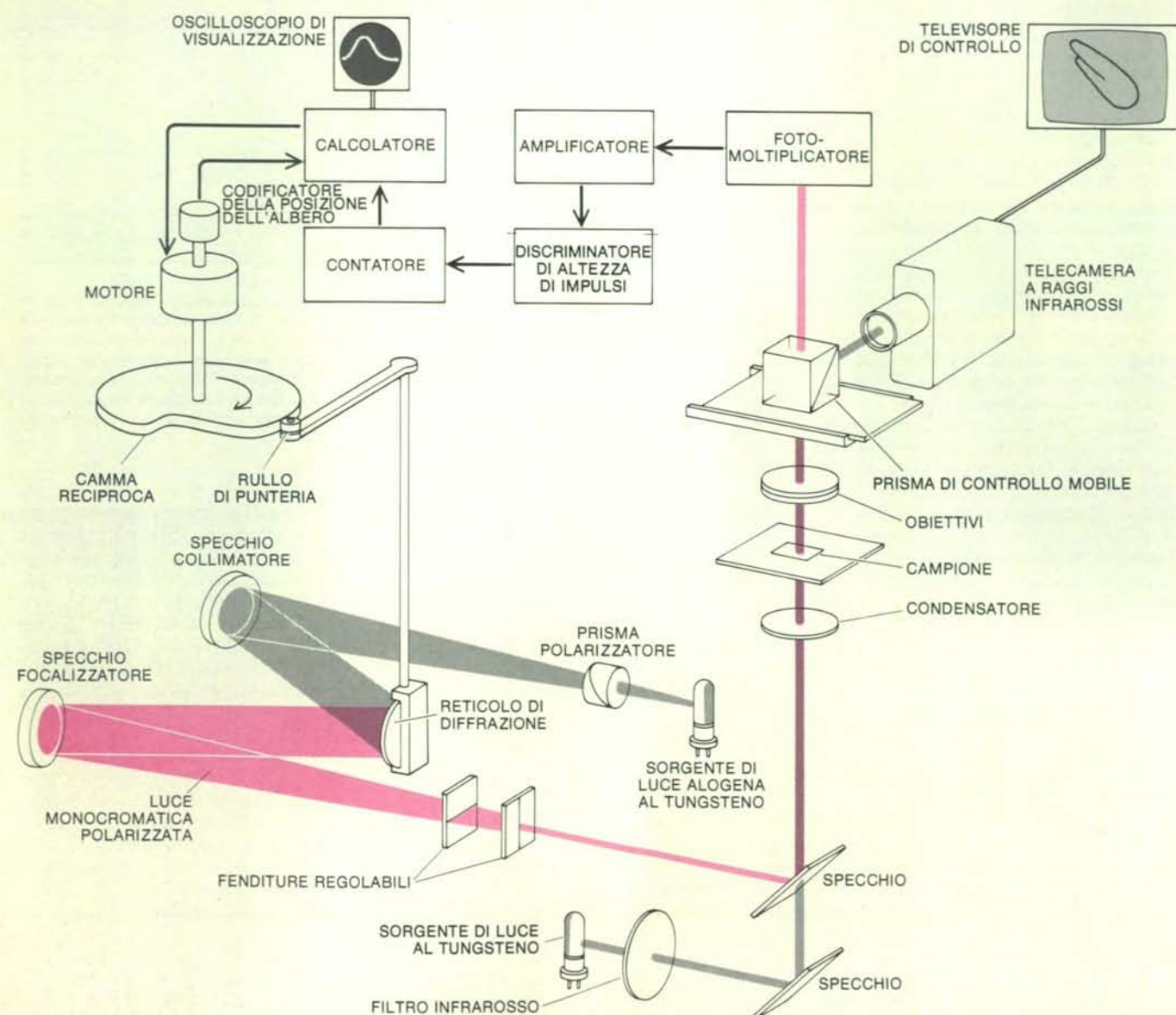
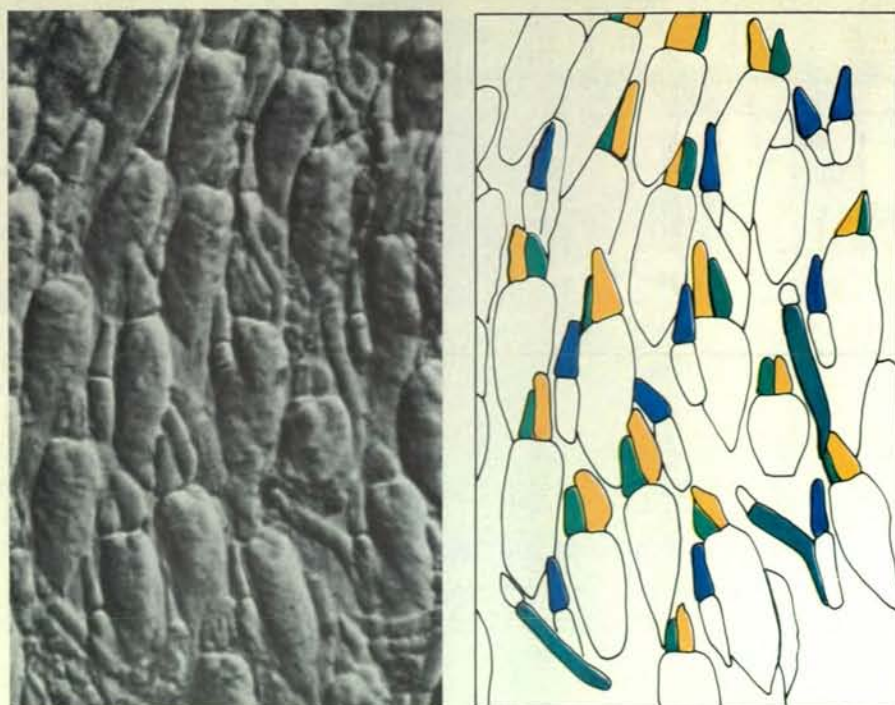
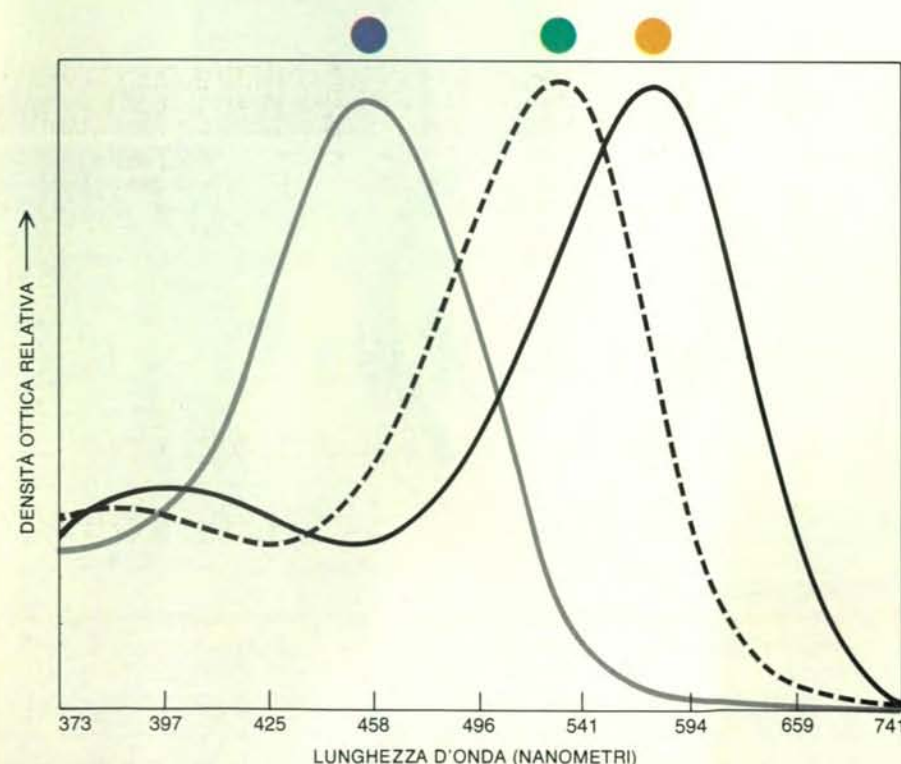
Nell'occhio di un pesce, come nell'occhio di un mammifero, i fotorecettori che raccolgono la luce entrante contengono sostanze fotosensibili chiamate pigmenti visivi. I pigmenti visivi furono scoperti indipendentemente nel 1877 dai fisiologi tedeschi Franz Boll e Wilhelm Friedrich Kühne. Boll notò che quando la retina di una rana viene rimossa dall'occhio inizialmente è di un rosso luminoso, poi scolorisce a contatto con la luce, diventando dapprima gialla poi incolore. Kühne scoprì subito dopo che anche il pigmento presente nei bastoncelli di molti animali è rosso e scolorisce quando viene esposto alla luce. In un animale vivente il colore del pigmento viene ripristinato se l'animale è posto al buio; ma nelle cellule prelevate dall'animale lo scolorimento del pigmento è permanente.

I pigmenti visivi catturano i fotoni di luce assorbendoli, proprio come fa l'acqua. Tuttavia per capire l'importanza dei pigmenti è necessario considerare quali sono le lunghezze d'onda maggiormente assorbite piuttosto che quelle trasmesse al massimo grado. Il pigmento nei bastoncelli umani, per esempio, è di colore rosa: da qui il nome di rodopsina (dal greco «rosa» e «visione»). La rodopsina appare rosa poiché trasmette luce rossa e blu, mentre presenta il massimo assorbimento alla lunghezza d'onda di 500 nanometri, nella regione blu-verde dello spettro.

Quando un pigmento visivo assorbe fotoni, le molecole del pigmento subiscono dei cambiamenti strutturali che innescano risposte elettrochimiche nei fotorecettori contenenti il pigmento. La risposta di ogni cellula è in parte determinata dal numero di fotoni assorbiti dal pigmento, ma non dalla lunghezza d'onda dei fotoni. Una volta che il fotone è stato assorbito, la sua lunghezza d'onda non influenza più il rendimento del fotorecettore. Il solo mezzo del sistema visivo per ricavare informazioni sul colore della luce incidente è di tener conto della probabilità statistica che uno dei suoi pigmenti assorba un fotone di una data lunghezza d'onda.

Come si possono determinare queste probabilità di assorbimento? I primi ricercatori estraevano chimicamente i pigmenti da un occhio intatto, di modo che le loro ricerche erano limitate quasi totalmente ai pigmenti dei bastoncelli. I pigmenti dei bastoncelli sono abbondanti e più facili da estrarre di quelli contenuti nei coni. Dopo l'estrazione del pigmento dal bastoncello, se ne esaminava lo spettro con uno spettrofotometro. In questo apparecchio il pigmento veniva attraversato da un raggio di luce monocromatica, con lunghezze d'onda selezionate nel campo del visibile, e veniva misurata l'intensità della luce trasmessa. L'intensità misurata veniva confrontata con l'intensità di un raggio di riferimento che non attraversava il pigmento, e che proveniva dalla stessa sorgente luminosa. È possibile trasferire i valori matematici ottenuti su un grafico che riporta la densità ottica del pigmento in funzione della lunghezza d'onda. Dal grafico, chiamato «spettro di assorbimento», si può inoltre calcolare la probabilità del pigmento visivo di assorbire un fotone di una determinata lunghezza d'onda.

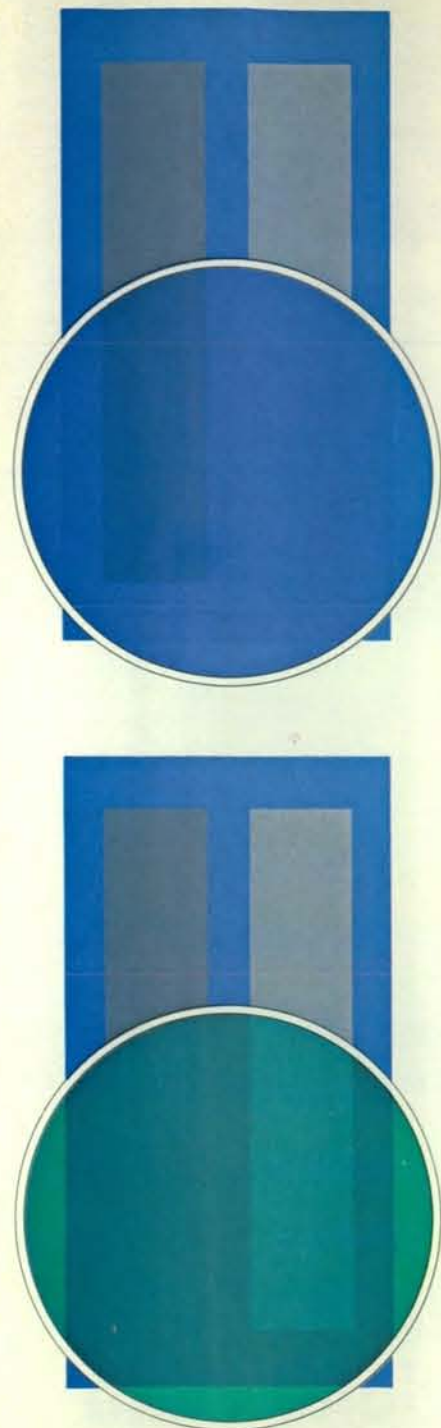
Il metodo che noi applichiamo per determinare gli spettri di assorbimento dei pigmenti visivi è concettualmente identico a quello sopra descritto; noi, però, siamo anche in grado di registrare gli spettri dei pigmenti presenti nei singoli fotorecettori. Lo strumento che usiamo è il microspettrofotometro, uno spettrofotometro che incorpora un microscopio con un elevato potere di risoluzione, cosicché è possibile analizzare lo spettro di una zona ingrandita dell'immagine del campione. Tali strumenti sono disponibili in meno di dodici laboratori in tutto il mondo. I primi tipi di microspettrofotometro del nostro laboratorio furono messi a punto da William B. Marks della Johns Hopkins University e da uno di noi (MacNichol) agli inizi degli anni sessanta, e da Ferenc I. Hárosi del National Institute of Neurological Diseases and Stroke agli inizi degli anni settanta. Tale strumento ci permise di esaminare oltre i vari pigmenti nei diversi tipi di fotorecettori, anche la distribuzione delle cellule nella retina.

La preparazione di un campione di retina per la microspettrofotometria consiste nel separare il sottile strato di fotorecettori dall'epitelio pigmentato, uno strato di tessuto retinico, posto dietro ai fotorecettori, che assorbe la luce deviata nell'occhio. Ciò si realizza immergendo dapprima la retina in una soluzione fisiologica, e separando quindi gli strati chirurgicamente. Lo strato di fotorecettori è posto su un sottile vetrino portaoggetti; sopra le cellule viene posto un secondo vetrino, in modo da appiattirle sul primo. Tutto questo lavoro deve essere fatto al buio o sotto luce infrarossa, altrimenti i pigmenti del campione potrebbero scolorirsi risultando inutili al fine di fornire dati relativi allo spettro. Quindi per posizionare il vetrino e mettere a fuoco l'immagine del campione da analizzare al microspettrofotometro noi utilizziamo luce infrarossa a una lunghezza d'onda che non viene assorbita in modo apprezzabile da alcun pigmento visivo. Osserviamo quindi il campione su un televisore di controllo collegato a una telecamera sensibile alla luce infrarossa che riceve un'immagine del preparato ingrandita mille volte.

**Il microspettrofotometro determina lo spettro di assorbimento di un pigmento visivo in un singolo fotorecettore. L'apparecchio misura il rapporto tra l'intensità di un raggio di luce monocromatica passante attraverso il pigmento e l'intensità di un identico raggio che non lo attraversa. Nel diagramma schematico ogni raggio è polarizzato e quindi scisso nei suoi componenti spettrali da un reticolo di diffrazione. La lunghezza d'onda del raggio monocromatico è calibrata sulla posizione di un albero a camme che controlla il reticolo. Quindi un calcolatore registra simultaneamente la precisa lunghezza d'onda del raggio incidente e l'intensità di quello uscente. I pigmenti, se esposti a luce viva, sbiancano e perdono la capacità di assorbire la luce. Perciò i raggi luminosi che passano attraverso i campioni sono mantenuti a bassa intensità, e i singoli fotoni nel raggio vengono convertiti in impulsi elettrici da un fotomoltiplicatore. Gli impulsi vengono allora amplificati, separati da segnali casuali e spuri, e registrati in un contatore, le cui risposte vengono lette periodicamente da un calcolatore. Per posizionare e focalizzare i campioni si usa un'illuminazione infrarossa in combinazione con una telecamera a raggi infrarossi e un televisore di controllo.**

I coni di *Gymnocorymbus ternetzii,* qui ingranditi 500 volte, differiscono nella forma e nel contenuto dei pigmenti. Si possono distinguere tre tipi di coni: grandi coni doppi, costituiti da cellule separate, che trasportano i pigmenti sensibili alla luce arancione e verde, e coni singoli più piccoli che trasportano il pigmento sensibile alla luce blu. Si possono vedere inoltre alcuni lunghi bastoncelli; nel *G. ternetzii* i bastoncelli sono sensibili soprattutto alla luce verde. La microfotografia è stata eseguita appiattendo le cellule mostrate nell'illustrazione di pagina 81, in alto. Le cellule devono essere mantenute tra due vetrini prima di tracciare gli spettri di assorbimento. Nel diagramma la posizione dei pigmenti delle cellule è indicata dai colori ai quali essi sono più sensibili.

Gli spettri di assorbimento mostrano, per ogni lunghezza d'onda della luce, la densità ottica relativa di un pigmento visivo, e possono essere correlati matematicamente alla probabilità che un fotone di quella lunghezza d'onda sia assorbito da un pigmento visivo. I colori alla sommità di ogni curva dei tre spettri di assorbimento di *Cichlasoma longimanus* indicano la comparsa nel sistema visivo umano di lunghezze d'onda alle quali ogni pigmento è maggiormente sensibile, a 455, 532 e 579 nanometri rispettivamente. L'assorbimento dei pigmenti è espresso in funzione della frequenza. Quindi, quando le stesse curve rappresentano l'assorbimento come funzione della lunghezza d'onda, gli incrementi della lunghezza d'onda variano lungo l'asse orizzontale.

Una volta che il fotorecettore è stato allineato sulla traiettoria del fascio a scansione del microspettrofotometro, la registrazione dello spettro di assorbimento è praticamente automatica. Un fascio di luce monocromatica di bassa intensità viene fatto passare attraverso l'apparecchio, dapprima senza il preparato cellulare, poi con il campione in situ. Un piccolo calcolatore registra e mette a confronto le intensità dei due fasci trasmessi per ogni lunghezza d'onda. Lavorando con fasci luminosi di bassa intensità e analizzando lo spettro rapidamente e ripetutamente, si evita la sostituzione per schiarimento del pigmento ogni volta che la luce passa attraverso le cellule del campione.

Quasi tutti i pigmenti visivi assorbono luce di tutte le lunghezze d'onda nel campo del visibile, in misura diversa. Tuttavia, per ogni pigmento esiste una lunghezza d'onda ben definita alla quale è massima la probabilità di assorbimento, e tali probabilità diminuiscono piuttosto bruscamente per lunghezze d'onda sia superiori sia inferiori. Un pesce che vive nelle acque profonde, con un solo pigmento visivo, avrebbe la più alta sensibilità visiva possibile se la massima probabilità di assorbimento del pigmento si accordasse con la luminosità di sfondo. Uno dei primi successi nel campo dell'ecologia visiva fu dato dalla conferma che nei pesci non migratori e abitanti le acque profonde marine queste concordanze sono comuni. A più di cento metri di profondità la luce disponibile è pallida e blu, a lunghezze d'onda comprese tra 470 e 490 nanometri. Le lunghezze d'onda relative al massimo assorbimento dei pigmenti dei bastoncelli sono in stretto accordo con la distribuzione sullo spettro della luce ambiente.

I pigmenti dei bastoncelli delle specie che abitano le acque costiere marine e le acque dolci non seguono, tuttavia, la distribuzione della luce disponibile sullo spettro così strettamente come avviene per le specie delle acque profonde. I bastoncelli dei pesci della zona temperata, il cui habitat è costituito da acque poco profonde e salate, sono più sensibili alla luce blu-verde con lunghezze d'onda comprese tra 500 e 510 nanometri, anche se la luminosità locale è giallo-verde a lunghezze d'onda comprese tra 525 e 550 nanometri. Sembra che i pigmenti bastoncellari nei pesci d'acqua dolce, il cui habitat varia da acque profonde a medioprofonde, abbiano la massima probabilità di assorbimento a lunghezze d'onda non superiori a 540 nanometri, anche se la luminosità di sfondo del loro ambiente è dominata da una luce rosso-arancio, con una lunghezza d'onda di 600 nanometri.

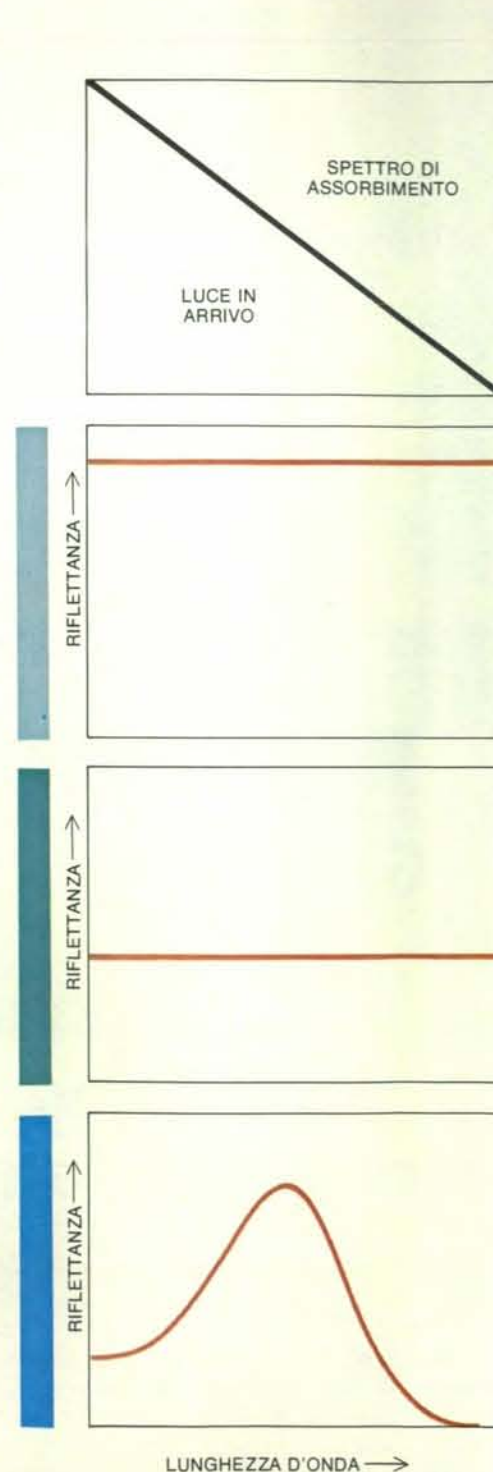
Nel 1966 J. N. Lythgoe dell'Università del Sussex suggerì che i pigmenti apparentemente disaccoppiati fossero controbilanciati dal colore della luce ambientale per aumentare il contrasto visivo. Lythgoe fece notare che quando un pigmento è in accordo con la luce ambientale esiste un contrasto relativamente alto tra un oggetto scuro, che deposita pochi fotoni sul pigmento visivo, e la luminosità

**Nei pesci i sistemi visivi a pigmenti multipli possono essersi inizialmente evoluti perché in questo modo aumentava il contrasto tra gli oggetti luminosi e scuri e la luce dell'ambiente, più che per una effettiva utilità nella discriminazione dei colori. Per un animale monocromatico il cui pigmento visivo sensibile al blu si accorda con la luce blu dell'ambiente, tipico degli habitat marini al largo delle coste *(illustrazione in alto)*, un oggetto scuro è chiaramente visibile sullo sfondo, mentre un oggetto luminoso risulta quasi invisibile. Un pigmento visivo sensibile al verde che non si accorda con lo sfondo *(illustrazione in basso)* determina un contrasto tra oggetto luminoso e sfondo, sebbene un oggetto scuro sia quasi invisibile. Pesci che posseggono entrambi i pigmenti possono distinguere oggetti scuri e luminosi nel loro ambiente, anche se magari mancano delle interconnessioni nervose necessarie per distinguere solamente in base al colore.**

di sfondo che ne deposita molti. Un oggetto luminoso, tuttavia, depositerà molti fotoni sul pigmento, indipendentemente dal suo colore; in questo modo esso non sarà distinguibile chiaramente dalla luminosità di sfondo. D'altra parte, se la luce che illumina l'oggetto ha un ampio spettro e il pigmento non è in accordo con la luce dell'ambiente, la situazione sarà ribaltata. Un oggetto scuro e la luce dell'ambiente appariranno entrambi scuri e saranno relativamente indistinguibili. Tuttavia, un oggetto luminoso apparirà luminoso e sarà così perfettamente distinguibile dalla luce dell'ambiente circostante *(si veda l'illustrazione a sinistra)*.

I bastoncelli sono sensibili in modo particolare alla luce di bassa intensità. Poiché in una data specie essi contengono normalmente solo un tipo di pigmento, la visione che deve dipendere dai bastoncelli è sensibile solamente alle variazioni di intensità luminosa. A un pesce che possiede solo bastoncelli (monocromatismo bastoncellare) il mondo appare come in un film in bianco e nero visto attraverso un filtro colorato. Oggetti blu scuro non sono distinguibili da quelli di colore rosso scuro o verde scuro e i colori brillanti vengono facilmente confusi col bianco.

Ciononostante un organismo può adattare il suo comportamento in modo tale da trarre il massimo vantaggio dal contrasto tra la luce dell'ambiente e gli oggetti relativamente scuri. Per esempio, sembra che la palamita *(Katsuwonus pelamis)* abbia solo un pigmento visivo, anche se possiede sia bastoncelli sia coni. Munz e McFarland notarono che la palamita compie evoluzioni che le consentono di attaccare la preda dal basso. Vista dal basso la preda si staglia contro la componente verticale della luce ambiente, e la palamita, il cui unico pigmento sensibile è in accordo con la luce dell'ambiente, può sfruttare il contrasto tra la sua preda relativamente scura e la luce dell'ambiente circostante relativamente luminoso.

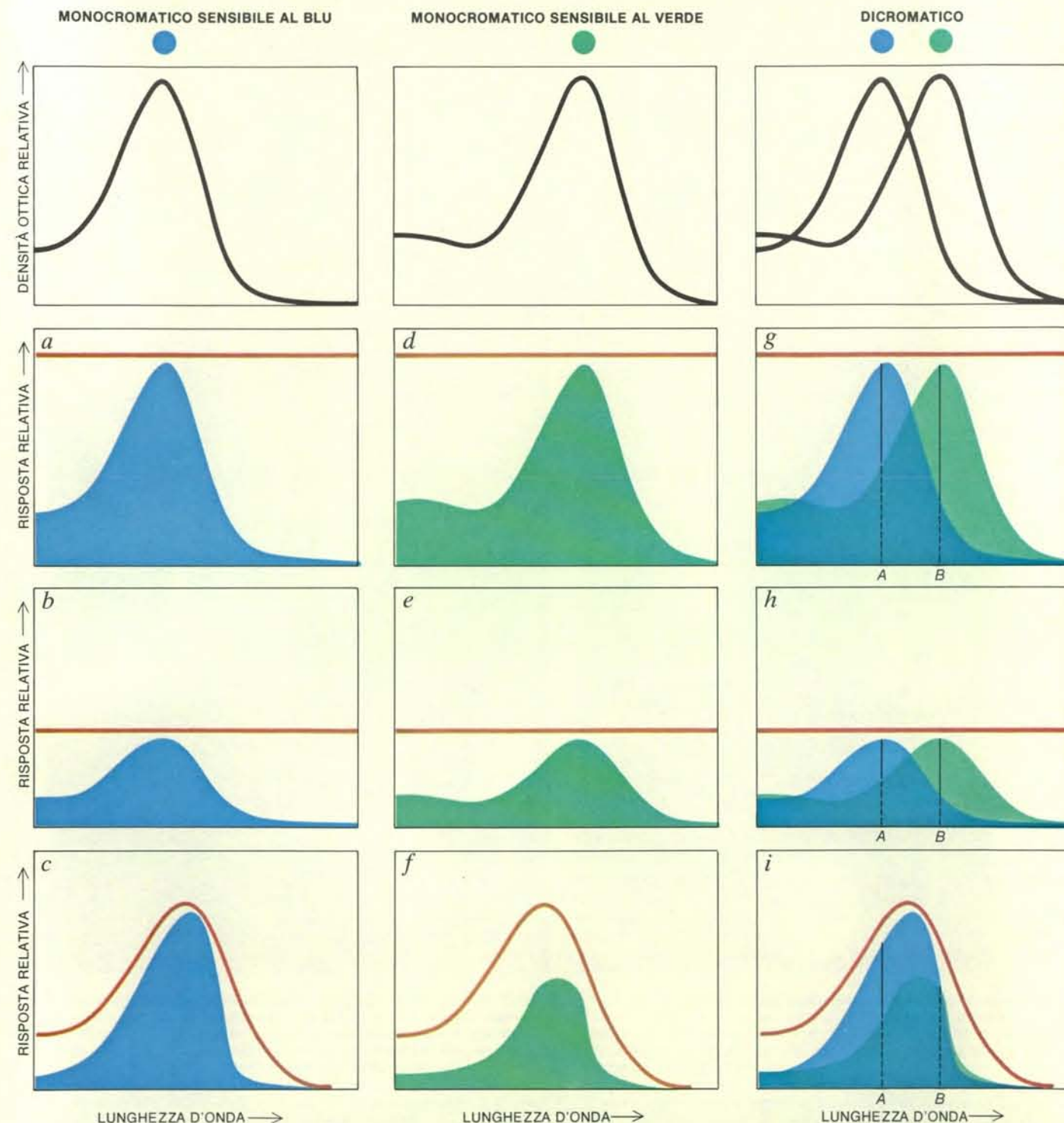
Molti pesci, tuttavia, posseggono due o tre classi di cellule a cono (o coni) oltre ai bastoncelli. Ogni classe di coni è associata di solito a un particolare pigmento visivo e misurazioni microspettrofotometriche hanno dimostrato che i pigmenti dei coni e dei bastoncelli hanno più o meno le stesse probabilità di assorbire fotoni. Ciononostante, a causa delle differenze tra bastoncelli e coni e della natura delle loro interazioni con le altre cellule della retina, l'attivazione della visione dei coni richiede una luce molto più intensa di quella necessaria per attivare la visione dei bastoncelli. I coni sono generalmente associati alla visione cromatica, e la capacità di distinguere i colori costituisce un chiaro vantaggio evolutivo in habitat in cui la luce è intensa e composta di molti colori. Tuttavia, i singoli coni non sono in grado di distinguere i colori meglio dei bastoncelli: anche la loro risposta dipende dal numero di fotoni assorbiti dai loro pigmenti e non dalle lunghezze d'onda dei fotoni. Anche quando nei fotorecettori della retina sono presenti parecchi tipi di pigmento, è necessario un ulteriore «cablaggio» nervoso perché si possano fare distinzioni sulla pura base del colore. Quindi è probabile che i vantaggi evolutivi conferiti inizialmente dai coni non fossero direttamente collegati alla visione colorata. Invece, i vantaggi offerti dalla presenza dei coni derivano probabilmente dal loro migliore adattamento alla luce

SPETTRO DI ASSORBIMENTO
LUCE IN ARRIVO
RIFLETTANZA
RIFLETTANZA
RIFLETTANZA
LUNGHEZZA D'ONDA

**La risposta di un fotorecettore a uno stimolo visivo dipende dall'intensità dello stimolo e dalle caratteristiche di filtro del pigmento assorbente. La lunghezza d'onda di un fotone in arrivo non influenza invece la risposta cellulare una volta che il fotone è stato assorbito.**

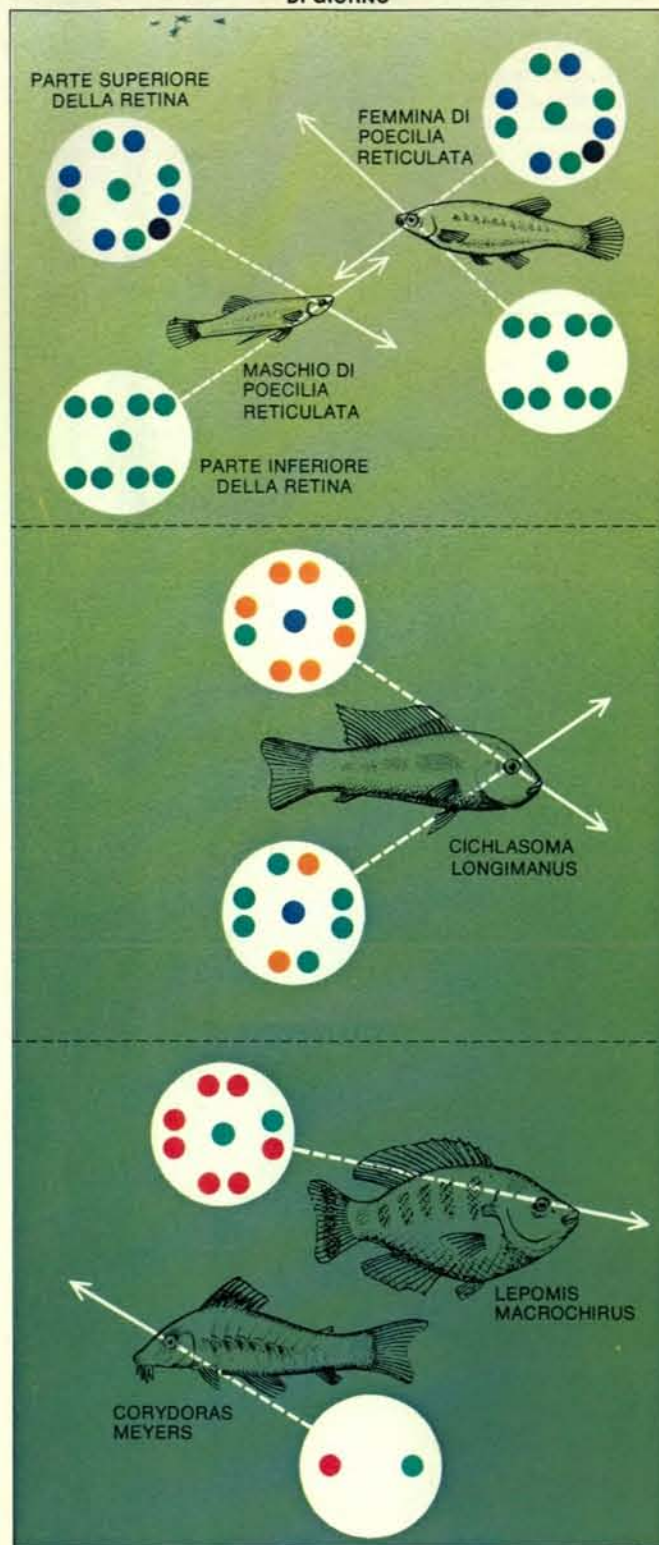
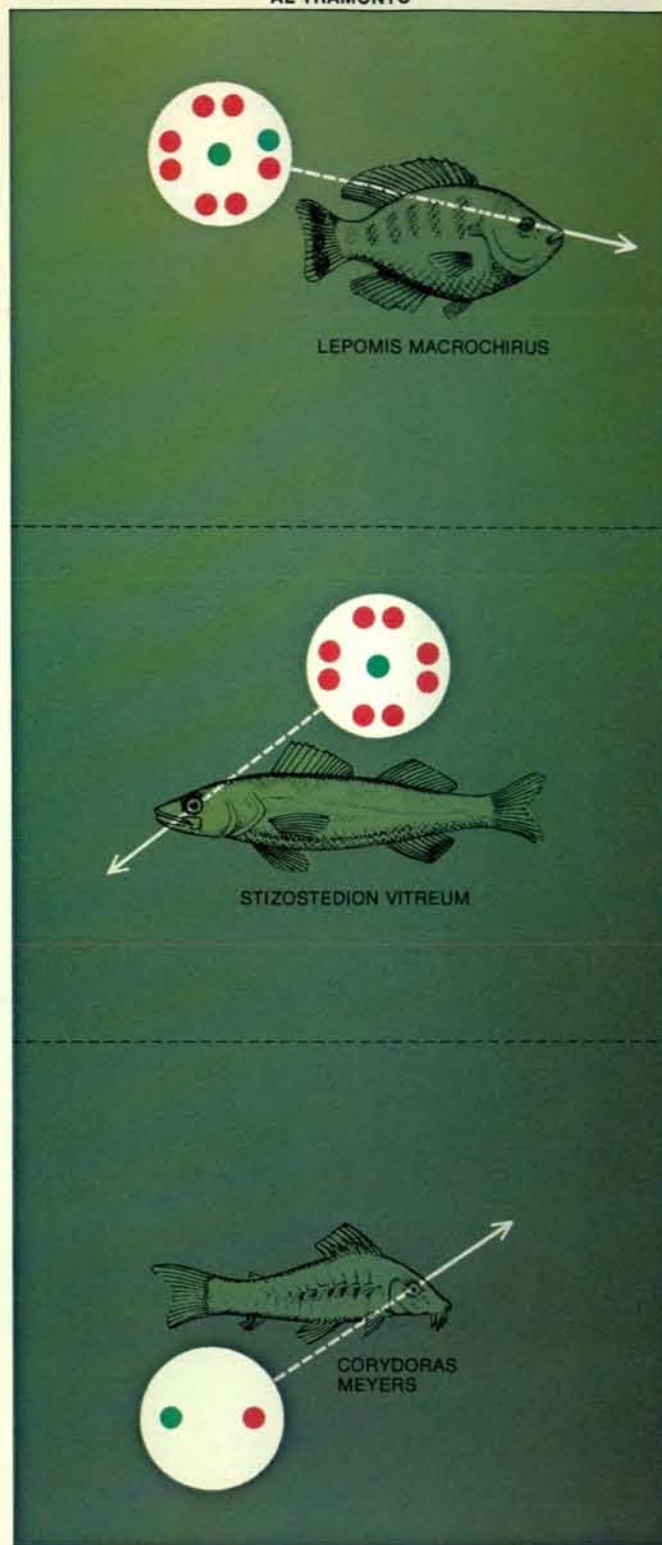


intensa e dal fatto che un sistema visivo composto da tanti pigmenti rende possibile il contrasto ottico su una banda dello spettro più ampia, a differenza di un sistema composto da un solo pigmento.

Alcuni pesci d'acqua salata che vivono alla superficie, come la corifena *(Coryphaena hippurus)*, hanno due differenti pigmenti nei coni: uno che assorbe soprattutto nel verde e uno che assorbe principalmente nel blu. Un tale sistema visivo dicromatico è sensibile alla luce attraverso una banda dello spettro più ampia rispetto a quella di un sistema monocromatico, semplicemente perché più pigmenti portano a un efficiente assorbimento in più regioni dello spettro. Quando un pesce dicromatico guarda verso l'alto il pigmento verde si trova in buon accordo con la luce dell'ambiente, mentre quando il pesce guarda in avanti o verso il basso è il pigmento blu che meglio si accorda con la luce dell'ambiente.

McFarland e Munz hanno, tuttavia, sostenuto che dal punto di vista evolutivo il principale vantaggio offerto da sistemi visivi a due pigmenti contro quelli a un

**La figura mostra come varia la risposta delle cellule *(aree in blu e verde)* con l'ingresso dei colori indicati nell'illustrazione a sinistra. La risposta dei coni sensibili al blu è simile a quella per la luce grigia e blu *(a, c)*. Le aree grigio scuro forniscono una risposta ridotta *(b)* e contrasterebbero così con lo sfondo per un monocromatico sensibile al blu. La risposta dei coni sensibili al verde, d'altra parte, è simile per i segnali in entrata grigio scuro e blu *(e, f)*, mentre la risposta per la luce grigia in arrivo *(d)* contrasta con lo sfondo. Un sistema di pigmenti dicromatico può anche essere la base della visione dei colori *(g, h, i)*. Una cellula che controllasse le differenze tra le risposte dei due tipi di coni sarebbe in grado di distinguere luce di lunghezza d'onda *A* da luce di lunghezza d'onda *B*, indipendentemente dall'intensità della luce.**

DI GIORNO

AL TRAMONTO

**La sensibilità dei pesci ai colori è specifica per le condizioni ambientali. Sono qui illustrati sei habitat visivi, relativi alle acque dolci di lago; i pesci sono ritratti nell'habitat in cui vivono. I cerchi colorati rappresentano gli schemi a mosaico dei fotorecettori in ogni specie; i colori corrispondono alla massima sensibilità per un colore di ogni pigmento visivo presente nelle cellule. La sensibilità al colore si sposta verso il rosso con l'aumentare della profondità e corrisponde a una crescente predominanza nell'acqua di luce ad elevata lunghezza d'onda. Due specie di pesci mostrano marcate differenze intraretinali nella distribuzione cellulare. La parte inferiore della retina, che raccoglie la luce proveniente dall'alto, del pecilide *Poecilia reticulata* è dominata dalla presenza dei coni sensibili al verde. Perciò il pecilide può distinguere oggetti scuri in una luce proveniente dal basso, ma la parte inferiore della retina non è in grado di distinguere i colori. La parte superiore della retina, che raccoglie la luce proveniente dal basso, è costituita da tutti e tre i tipi di coni ed è perciò sensibile a un vasto spettro di colori. Durante l'accoppiamento il maschio di pecilide si porta al di sotto o di fronte alla femmina in modo che i suoi colori possano essere registrati nella zona più sensibile ai colori della retina della femmina. La parte inferiore della retina del ciclide, che raccoglie la luce proveniente dall'alto, è sensibile essenzialmente alla radiazione luminosa di media lunghezza d'onda, mentre la parte superiore della retina, che raccoglie la luce proveniente dal basso, è più sensibile alla luce di elevata lunghezza d'onda. Le differenze corrispondono alla distribuzione sullo spettro della luminostià di sfondo lungo le differenti linee di vista, e alla diversità dei compiti, per i quali è indispensabile la vista, tipici di ogni specie. Gli schemi a mosaico sulla retina di *Lepomis macrochirus, Corydoras meyers* e di *Stizostedion vitreum* sono tutti abbastanza uniformi. *Lepomis macrochirus* si ciba di fitoplancton vicino alla superficie, al tramonto o all'alba, e insegue il plancton verso il fondo per tutto il giorno.**

pigmento risiede nel fatto che sistemi a due pigmenti permettono di distinguere sulla luminosità di sfondo sia oggetti scuri sia oggetti luminosi. Quando un animale dicromatico sensibile alla luce blu e verde guarda verso l'alto, il pigmento che assorbe la luce verde consente al pesce di distinguere gli oggetti scuri contro la luce ambientale prevalentemente verde. Il pigmento che assorbe la luce blu permette la differenziazione degli oggetti luminosi. Quando il pesce guarda in avanti o verso il basso, i ruoli dei due pigmenti vengono ribaltati, ma il risultato complessivo è lo stesso: sia gli oggetti luminosi sia quelli scuri possono essere distinti sullo sfondo.

I dati da noi ottenuti dimostrano che i pigmenti dei coni dei teleostei si adeguano piuttosto bene alla quantità di luce disponibile nell'ambiente dei pesci. La gamma più ampia di pigmenti visivi si trova tra le specie che vivono vicino alla superficie, dove la luce è intensa e ricopre una vasta zona dello spettro. Nell'acqua salata a medie profondità la luce rossa di elevata lunghezza d'onda, proveniente dalla superficie, viene assorbita quasi interamente e a tali profondità la maggior parte delle specie mancano dei pigmenti che assorbono la luce rossa. I loro coni sono sensibili in modo particolare alla luce blu e verde.

Un'eccezione a questa tendenza è costituita da *Hemitripterus americanus* che noi studiammo con Hárosi e Barbara-Ann Collins del Marine Biological Laboratory a Woods Hole, nel Massachusetts. *H. americanus* possiede tre pigmenti sensibili in modo particolare alla luce blu, verde e giallo-verde che coprono una zona dello spettro molto più ampia di quanto ci si aspetterebbe per un pesce che passa la maggior parte della sua vita a profondità superiori a cento metri. Tuttavia, per comprendere l'utilità del sistema visivo di un pesce se ne devono considerare tutti gli aspetti comportamentali. Nella stagione degli accoppiamenti, in autunno avanzato, il pesce si muove verso l'acqua bassa e quindi verso una luce più brillante, che interessa uno spettro più ampio. Il corpo degli esemplari che si riproducono varia fra giallo luminoso, arancio e scarlatto, e le uova fecondate vengono depositate su una specie di spugna dal colore arancio brillante o giallo. Per tale comportamento è necessario saper distinguere nelle regioni dello spettro delle medie ed elevate lunghezze d'onda, ed è quindi indispensabile una vasta gamma di pigmenti visivi.

Nella maggior parte degli habitat di acqua dolce l'utilità di coni sensibili alla luce blu diminuisce con la profondità, poiché la luce trasmessa dalla superficie tende al rosso invece che al blu. I sistemi visivi dei pesci di acqua dolce che abbiamo esaminato possono essere divisi in quattro gruppi, in funzione della sensibilità di massimo assorbimento dei loro pigmenti visivi e della morfologia dei loro fotorecettori. Al primo gruppo appartengono pesci come il pecilide *Poecilia reticulata* e il danio gigante *(Danio malabari-*

*cus)*, che vivono nella zona immediatamente sotto la superficie dell'acqua. I loro sistemi visivi sono simili a quelli delle specie che vivono nelle acque marine poco profonde; sono caratterizzati da tre pigmenti visivi, sensibili soprattutto al violetto, al blu-verde e al giallo-verde.

La sensibilità della retina di *Poecilia reticulata* è strettamente legata ai suoi schemi comportamentali. La zona superiore della retina, sulla quale converge la luce proveniente dal campo visivo posto al di sotto e davanti a *Poecilia*, è un mosaico di coni sensibili a tre regioni dello spettro ampiamente intervallate; questa zona è ben strutturata per la visione dei colori. La zona inferiore della retina, invece, sulla quale converge la luce proveniente dall'alto del campo visivo, è rivestita da cellule contenenti lo stesso pigmento sensibile al verde. John A. Endler dell'Università dello Utah ha recentemente sottolineato che quando maschi di *Poecilia*, per attrarre le femmine, mettono in mostra la loro intensa colorazione lo fanno da una posizione più avanzata e leggermente al di sotto dell'oggetto del loro desiderio. Tale posizione, nel campo visivo della femmina, assicura che l'immagine del maschio cada nella regione maggiormente sensibile ai colori della retina della femmina. La zona inferiore della retina sensibile al verde è adatta per la ricerca di pezzi di cibo scuri che si stagliano sulla componente verticale della luce ambiente, principalmente verde.

Anche il secondo gruppo di pesci d'acqua dolce possiede tre tipi di pigmento visivo, ma le loro sensibilità massime risiedono nelle regioni dello spettro con lunghezze d'onda superiori a quelle dei pigmenti del primo gruppo. Tali pesci includono un ciprinide *(Barbus tetrazona)* e un ciclide *(Cichlasoma longimanus)*. La loro sensibilità alla luce di elevata lunghezza d'onda è una chiara risposta evolutiva alla rapida scomparsa della luce di bassa lunghezza d'onda nelle acque molto assorbenti che essi abitano.

Il lepomide *(Lepomis macrochirus)*, l'occhiogrigio *(Stizostedion vitreum)*, il piranha *(Serrasalmus sp.)* e altri pesci appartenenti al terzo gruppo hanno poche se non solo qualche cellula sensibile al blu, nelle loro retine, sebbene abbiano cellule sensibili alla luce verde e rossa. Essi sono attivi vicino alla superficie, all'alba o all'imbrunire, anche se molte specie trovano nutrimento nelle acque più profonde durante la giornata. Quindi il loro ambiente spettrale rimane più o meno lo stesso durante il loro periodo attivo. I pigmenti di questi pesci sensibili alla luce rossa fanno sì che captino meglio dell'uomo la radiazione avente una elevata lunghezza d'onda. La cosiddetta acqua nera abitata da molti di tali pesci è in realtà illuminata da una luce ambiente nell'infrarosso vicino, che essi possono vedere.

Le retine delle specie appartenenti al quarto gruppo sono costituite interamente da bastoncelli e da coni sensibili alla luce verde e rossa. In questi coni non sono stati trovati pigmenti sensibili alla luce blu e sembra che alcune specie manchino anche dei pigmenti sensibili alla luce verde. Probabilmente i pesci conservano i coni sensibili al rosso solo perché la luminosità di sfondo tende a spostarsi verso il rosso; altrimenti sarebbero con ogni probabilità affetti da acromatopsia bastoncellare. Le specie appartenenti a questo gruppo includono numerosi pesci «spazzini» come *Corydoras meyers* e il labeo bicolore *(Labeo bicolor)*, dalla caratteristica coda rossa.

È probabile che la visione dei colori, la capacità di distinguere la luce in base sia alla lunghezza d'onda sia all'intensità, si sia evoluta qualche tempo dopo i sistemi visivi a due pigmenti. Per essere in grado di distinguere i colori, un pesce deve possedere non soltanto due o più pigmenti visivi i cui spettri di assorbimento coincidano; esso deve anche disporre di interconnessioni nervose in grado di confrontare in maniera appropriata i segnali provenienti dai fotorecettori. Per capire come ciò si realizza, consideriamo l'effetto di uno stimolo rosso monocromatico avente lunghezza d'onda di 630 nanometri sull'occhio di un ciclide. L'occhio del ciclide possiede tre pigmenti visivi incluso quello che assorbe la luce gialla a una lunghezza d'onda di 579 nanometri e quello che assorbe la luce verde a 532 nanometri. Indipendentemente dall'intensità dello stimolo rosso, il pigmento che assorbe la luce gialla assorbirà più fotoni rossi del pigmento che assorbe la luce verde. D'altra parte, un impulso giallo-verde, con lunghezza d'onda di 540 nanometri, avrà più o meno lo stesso effetto sul pigmento che assorbe la luce gialla e su quello che assorbe la luce verde, indipendentemente dall'intensità luminosa dello stimolo.

Il rendimento di una cellula nervosa che controlla entrambi i tipi di fotorecettori può dipendere dalla differenza dei due segnali in entrata. Una tale cellula permetterebbe al ciclide di distinguere il rosso dal giallo-verde, anche se la luminosità dei due colori, così come viene misurata da una classe di coni del ciclide, potrebbe essere identica. L'esistenza di cellule che elaborano i segnali provenienti dai fotorecettori di differenti classi fu scoperta per la prima volta nella retina dei pesci dal fisiologo svedese Gunnar Svaetichin e da uno di noi (MacNichol) negli anni cinquanta. Esse sono chiamate cellule antagoniste del colore e sono state trovate a parecchi livelli del sistema visivo di molte specie.

Poiché le cellule antagoniste del colore dipendono dalla differenza tra i segnali dei fotorecettori, la differenziazione del colore è più accurata nelle regioni dello spettro in cui gli spettri di assorbimento di almeno due pigmenti visivi si sovrappongono e mostrano pendenze marcatamente differenti. Se gli spettri di assorbimento hanno la stessa pendenza o non si sovrappongono, la differenza di rendimento delle due classi associate di fotorecettori è una costante e la differenziazione del colore è minima. L'aggiunta di un terzo o di un quarto pigmento al sistema visivo, insieme ad adatte cellule antagoniste del colore, può rendere possibile la distinzione del colore in queste regioni «morte» dello spettro. Naturalmente un terzo o quarto pigmento può anche favorire la completa sensibilità di un sistema visivo persino in una banda di lunghezze d'onda più ampia di quanto possano fare due pigmenti.

Considerati i vantaggi evolutivi offerti dalla capacità di distinguere il colore oltre alla pura e semplice capacità di distinguere l'intensità delle luce, perché la visione dei colori non è comune a tutti i teleostei? Perché funzioni bene, una retina sensibile ai colori deve essere immersa in luce sufficiente a coprire un'ampia banda di lunghezze d'onda al fine di stimolare segnali consistenti da tutti i tipi di coni. Tutti i fotorecettori che non emettono segnali lasciano dei «buchi» nell'immagine visiva. Tali buchi degradano l'immagine, e così i pigmenti dei coni nella retina che non contribuiscono regolarmente all'immagine visiva sono in realtà non funzionali. Un pesce attivo solo di notte o al tramonto, o che vive in acque marine profonde, raramente incontrerà luce sufficiente per stimolare completamente i coni. Per un tale pesce, e apparentemente per molti primati parenti dell'uomo, il prezzo della sopravvivenza al buio è la perdita della sensibilità al colore.

# Leonardo Pisano

*La lettura del «Libro dei quadrati» di Leonardo Pisano conferma, contro l'opinione di vari critici moderni, la correttezza e l'originalità del più grande matematico del Medioevo, noto anche con il nome di Fibonacci*

di Ettore Picutti

I traffici con l'Oriente iniziati già prima del Mille dalle Repubbliche marinare italiane e la penetrazione dopo il Mille in territori di cultura araba dei normanni di Sicilia, della Reconquista spagnola e dei crociati, portarono alla rinascita della cultura europea del dodicesimo secolo; tale cultura sorgeva con un'impronta greco-araba, filosofico-scientifica, sulla quale l'impronta latina letteraria si sarebbe sovrapposta un secolo dopo.

Ingrediente primo della rinascita fu l'entusiasmo con il quale studiosi laici ed ecclesiastici di tutte le parti d'Europa si misero alla ricerca dei documenti dell'antichità greca tradotti in arabo e di originali arabi, entusiasmo adombrato dalle tradizioni sul viaggio nella Spagna musulmana di Gerberto d'Aurillac e sulla conversione all'Islamismo per amor di sapere di Adelardo di Bath, e attestato dalla presenza di italiani, inglesi, francesi, tedeschi tra i traduttori dall'arabo della Scuola di Toledo.

La matematica prendeva avvio deciso dalla traduzione in latino degli *Elementi* di Euclide (Adelardo di Bath e Gherardo da Cremona), delle opere di aritmetica e algebra del persiano al-Khuwarizmi degli inizi del IX secolo (Adelardo di Bath, Roberto di Chester), del *De mensura circuli* di Archimede (Gherardo da Cremona, Platone da Tivoli), del *Liber trium fratrum* di geometria greco-araba del IX secolo (Gherardo da Cremona). Essa rinasceva con aspetto nuovo, quasi anti-greca nel suo spirito, in quanto non fine a se stessa e godimento spirituale per l'*otium* del filosofo, ma deliberatamente applicativa quale i tempi nuovi la esigevano.

In questa atmosfera di concretezza intellettuale si formò alla matematica Leonardo Pisano dei figli di Bonaccio a fine secolo. Erano i tempi delle imprese del Saladino e di Riccardo Cuor di Leone; mentre risuonava l'eco di queste imprese i mercanti pisani, genovesi e veneziani continuavano a estendere i loro fondachi nei porti del Mediterraneo e del Mar Nero.

Leonardo nacque attorno al 1170 ed era quindi coetano di S. Domenico e di una decina d'anni più anziano di S. Francesco. Abbiamo informazioni sugli inizi della sua carriera matematica dalla prefazione alla sua prima opera, il *Liber abaci* scritto nel 1202.

Quand'era ancora ragazzino, suo padre, incaricato di dirigere per conto dell'*Ordo Mercatorum* di Pisa l'ufficio doganale di Bugea in Algeria, lo chiamò presso di sé e gli fece seguire per parecchi giorni un corso sul calcolo posizionale indiano, del quale a un esperto non potevano sfuggire i pregi. Fu questo il suo avvio alla matematica; nei suoi frequenti viaggi per motivi di lavoro, certo sempre per conto dei mercanti pisani, approfittò poi per far conoscenza con i matematici delle regioni visitate, Egitto, Siria, Provenza, Grecia, per intrattenersi con loro in discussioni e sfide *(disputationis didici conflictum)* e per studiarsi a fondo gli *Elementi*, che tenne poi sempre a modello di rigore logico e di stile.

Nacque così, tra un contratto e un controllo contabile, tra rollii e beccheggi di una galera pisana, il *Liber abaci*, primo e insuperato modello di «summa» matematica medioevale, con lo scopo dichiarato da parte del suo autore di mettere tutto il suo sapere in fatto di aritmetica e algebra «a disposizione della gens latina in modo che ben poco sull'argomento potesse restarne fuori».

Titolo sbagliato, commenta paradossalmente Carl Boyer nella sua *Storia della matematica*, ricordandosi forse che l'abaco per greci e romani e per i maestri d'abaco dei secoli precedenti Leonardo Pisano, a pedine o a gettoni che fosse, era uno strumento di calcolo. Leonardo invece riserva la denominazione di «abaco» all'aritmetica-algebra applicata in genere; era certo questo il significato dato al termine ai suoi tempi, significato che si mantenne in Italia sino al Settecento inoltrato.

Si tratta di un'opera colossale (459 pagine nell'edizione in 4° del Boncompagni) nella quale vengono presentate le «*novem figure*» degli indiani e il «*signum*» 0 *(quod arabice zephirum appellatur)*, le operazioni con esse in interi e in frazioni, le prove per 7, 9, 11, 13 e il criterio di divisibilità per 9, le applicazioni per determinare massimo comun divisore e minimo comune multiplo; vi si danno poi, corredate da numerosi problemi, regole su compra-vendite, baratti, società, leghe e cambi con le monete più varie allora in uso, proporzioni, tre semplice e composto e così via. Seguono capitoli a parte per la regola «*elchataym*» (di doppia posizione per la soluzione delle equazioni) e per le questioni «*aliebre et almucabale*» sino a soluzione, discussione, applicazione delle equazioni di secondo grado.

Sul *Liber abaci*, per oltre tre secoli, sino a Pacioli, si formeranno maestri ed allievi della scuola toscana; l'equilibrio teoria-pratica vi era curato e raggiunto («ho dimostrato con prove certe quasi tutto quello che ho trattato»).

Non era e non è un'opera facile, e Leonardo Pisano consigliava il lettore a insistere nelle applicazioni con esercizio continuo sino a che memoria e ragionamento, mani e numeri «quasi uno impulsu et anelitu in uno et eodem instanti circa idem per omnia naturaliter consonent». Sarà questo anelito di perfezione a fare di Leonardo un matematico di eccezione tra contemporanei e successori, i quali conservarono un reverente ricordo del maestro.

Commentava Antonio de' Mazzinghi nel secolo XIV: «O L.p. di quanta scientia fusti, quando desti principio all'Italia ad aver lume della praticha d'arismetricha!» (dal Codice Ottoboniano 3307 della Biblioteca Vaticana).

Maestro Benedetto da Firenze a metà del secolo successivo lo tradurrà, citerà ripetutamente le sue opere e ne ricorderà la figura: «Dicho che L.p. fu uomo sottilissimo in tutte le dispute et sechondo che si truova, lui fu il primo che ridusse allume questa praticha in toschana, che allora s'andava per vie molte e strane». E ancora: «El trattato de' numeri quadrati è la [sic] più difficile... E no ò trovato chi più alta mente ne dicha che L.p.» (Dal Codice L.IV.21 della Biblioteca civica di Siena).

Boyer troverà la lettura del *Liber abaci* «non molto appassionante per il lettore moderno», rimprovererà a Leonardo Pisano di aver usato la complicata forma di espressione algebrica dei suoi tempi, e sia a lui che a un millennio di matematici di non aver applicato le frazioni decimali. Leonardo, diciamo di conseguenza, passò la fiaccola a Viète e a Stevino, che vi giunsero quattro secoli dopo e proseguì per la sua strada; anzi, dovendo una volta estrarre la radice quadrata di 7234, aggiunse *quatuor zephira*, trovò la radice 8505 di 72340000, poi ripassò i decimali a frazione normale 1/20.

Dal 1202 al 1220 Leonardo Pisano non scrisse più nulla. Furono vent'anni fitti di storia e di formazione culturale per la civiltà europea.

Gli scomunicati della quarta crociata fondavano l'impero latino d'Oriente, e nuovi testi, greci questa volta, passavano in Europa; crociati comunicati distruggevano la Provenza e i suoi abitanti, demandando a Dio l'incarico di riconoscere le anime dei non eretici. A Parigi veniva proibita, pena la scomunica, la lettura in pubblico e in privato delle opere scientifiche di Aristotele. Il 1212 segna poi la fine degli Almohadi di Spagna; due anni dopo la corona inglese perderà i suoi possedimenti in Francia e dovrà concedere la Magna Charta. S. Francesco parlava agli uccelli e parlava al sole, alla luna e alle stelle, che i filosofi stavano incapsulando nel mostruoso meccanismo cosmico di Aristotele.

All'orizzonte della storia e della cultura italiana ed europea, si profilava allora la figura di Federico di Svevia «stupor mundi» con la sua Corte di notari e protonotari indigeni e «magistri» e «philosophi» di tutte le nazioni.

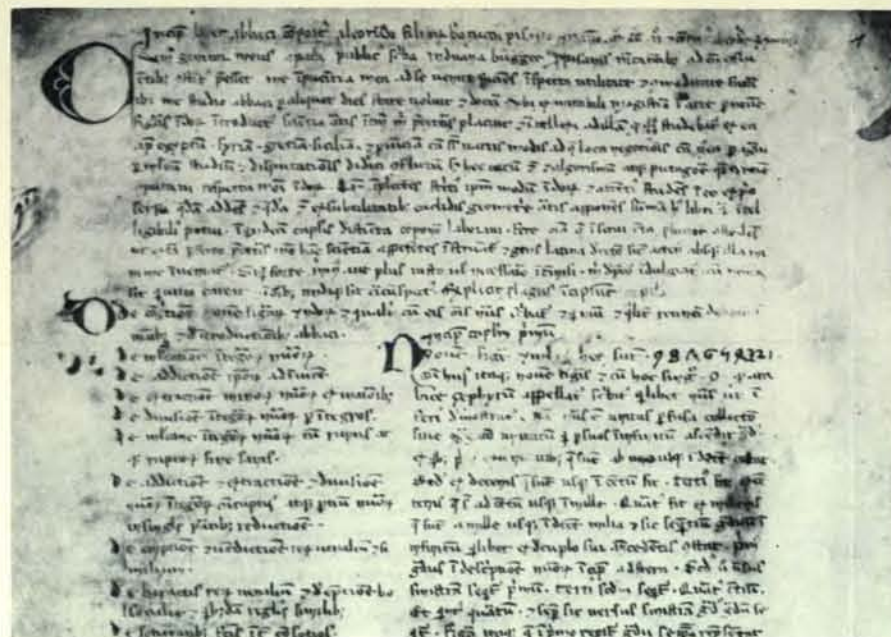
Leonardo Pisano avrebbe forse iniziata e conclusa con il *Liber abaci* la sua attività di matematico, senza l'intervento proprio di uno dei filosofi della Corte di Federico, Maestro Domenico, che egli chiamava amico e che da amico effettivamente si comportò, sia spingendolo a comporre la sua seconda «summa», la *Pratica geometrie (Practica geometriae*, nell'edizione del Boncompagni), sia interessandosi a presentarlo, qualche anno dopo, all'imperatore.

Nel 1220 l'opera era compiuta; essa consta di 223 pagine nell'edizione del Boncompagni.

La *Pratica geometrie* è certamente meno originale e varia del *Liber abaci*, ma si presenta comunque come un *corpus* di valore didattico eccezionale, anche per una scuola media superiore moderna. Scopo dichiarato dell'autore era quello di preparare un *perfectum documentum*, valido sia per gli appassionati delle *subtilitates* che per i pratici. E non si può dire che non lo abbia raggiunto.

Si basa sugli *Elementi* di Euclide (e sul suo libro perduto *Sulla divisione delle figure*), ma certe parti sono prese da Archimede, da Erone, da Tolomeo, da Savasorda (l'ebreo spagnolo bar Hiyyia, «sa-

**Miniatura al primo foglio del Cod. Urb. Lat. 292 della Biblioteca Apostolica Vaticana, dal quale Baldassarre Boncompagni ha tratto nell'Ottocento la sua edizione della *Practica geometriae*. È un evidente tentativo di raffigurare Leonardo Pisano nell'atto di presentare la sua opera.**

**Caratteri usati dai maestri d'abaco. Leonardo Pisano divulgò le cifre indiane e il loro uso. Sin dal decimo secolo, però, i maestri d'abaco usavano, per calcolare sulle loro tavolette, gettoni sui quali erano riportati segni analoghi alle cifre indiane. I maestri d'abaco indicavano questi segni con termini di probabile origine araba, ma ne sostenevano una origine pitagorica.**

*L'Incipit* del *Liber abaci* nel Cod. I 72 Sup., del XIII secolo, della Biblioteca Ambrosiana di Milano. Leonardo vi descrive il suo avvio alla matematica e vi presenta le cifre indiane. Rispetto alla parte iniziale del Cod. Magliabechiano CI 2616 di Firenze, dal quale il Boncompagni trasse la sua edizione, manca la dedicatoria a Michel Scotto per la seconda stesura del 1228.

Probabile busto di Federico II conservato nel Museo di Barletta; in realtà sembra che la sua figura avesse lineamenti meno classici. Gli arabi che lo conobbero durante la crociata del 1228-1229 ebbero alta considerazione delle sue qualità, ma non furono altrettanto colpiti dal suo aspetto. Sibt ibn al Giawzi diceva di lui che «come schiavo non sarebbe stato valutato duecento dirham».

hib al surta», cioè capo delle guardie, autore del *Liber Embadorum* tradotto dall'ebraico in latino da Platone di Tivoli agli inizi del XII secolo), nonché dai «Tre fratelli».

Le sue fonti sono quindi essenzialmente greche (né in fatto di geometria poteva essere diversamente); ma per i suoi scopi anche applicativi e per la ricerca di soluzioni alternative il suo spirito non è certamente greco.

Quanto al contenuto si può dire in breve che vi viene spiegata e applicata la parte sostanziale degli *Elementi*, inclusi i libri XIV e XV, premesse le nozioni algebriche necessarie alla soluzione dei problemi. Vi si trovano dimostrati in vari modi il teorema di Erone e quello di Tolomeo, e vi viene ricavato «con numeri piccoli» il valore di $\pi$.

La estesa e preziosa *Distinctio quarta* tratta della suddivisione delle figure. Per la valutazione delle aree dei campi vi viene anche esposto l'uso dell'archipenzolo e di un quadrante graduato con lancetta mobile, mediante il quale si possono determinare funzioni trigonometriche.

La *Pratica geometrie* del 1220 era l'indiretto omaggio del matematico pisano a Federico di Svevia che a fine anno, a 26 anni d'età, cingeva la corona imperiale, e che si sarebbe rivelato come il più colto e organizzato degli imperatori germanici.

Come il *Liber abaci*, così la *Pratica geometrie* diventava documento basilare per i maestri della scuola toscana, da Paolo dell'Abbaco a Cristofano di Gherardo, a Maestro Benedetto a Luca Pacioli.

Mentre la *Pratica geometrie* veniva completata, l'immane sciagura della quale, diceva Ibn al Athir, i giorni e le notti non avevano mai visto l'uguale dalla creazione del mondo, si abbatteva da Oriente sul mondo arabo. Gengiz Khan irrompeva nel Khowarezm e in Persia, e in due anni vi distruggeva per sempre secoli di cultura.

Nessuna delle numerose opere composte in tre secoli dai matematici toscani ebbe l'onore della stampa in periodo rinascimentale; il ricordo di un'attività eccezionale e unica in Europa in parte andò disperso, in parte rimase sepolto nel fondo delle biblioteche. Tutto restò affidato a quanto riportato nella *Summa de Arithmetica geometria. Proportioni et proportionalità* di frate Luca Pacioli, stampata a Venezia nel 1494 e ristampata nel 1523 «in Toscolano su la riva del Benacense et unico carpionista Laco» per i tipi di Paganino de' Paganini di Brescia.

Così Pacioli introduceva la sua opera: «E queste cose con tutte le seguenti: saranno sechondo gli antichi... e de' nostri moderni Leonardo Pisano. Giordano. Biagio da Parma. Giovan sacrobusco e Prosdocimo padoano/ da i quali in magior parte cavo el presente volume.»

Alla *Summa* restava affidato in particolare il ricordo di un'opera per altro sconosciuta di Leonardo Pisano. Dando la soluzione del problema di scomporre un numero in somma di due quadrati, così scriveva frate Luca: «... mai falla questa regola. La quale donde proceda Leo. Pi. nel tractato che fa de' quadratis numeris la dimostra per via de figure geometriche».

A quest'opera di Leonardo Pisano sui numeri quadrati Pacioli sembrava essersi particolarmente interessato; egli insisteva soprattutto su un problema in essa considerato: «Trovame un numero quadrato che trattone certa quantità remanghi quadrato: e giontoci la medesima quantità ancora sia quadrato. E simili a questa. Le quali domande sonno difficilissime quanto ala dimostrazione dela pratica: Commo sa chi ben la scrutinato. Maxime Leonardo pisano in un particulare tractato che fa de quadratis numeris intitulato. Dove con grande sforzo se ingegna dare norma e regola a simili solutioni. E pur finalmente generaliter non servano a tutte: e pur si convien a tastoni redurse in cercarle... Hoc opus hic labor est.»

Tale problema è traducibile in forma algebrica nella doppia equazione:

$$Y^2 - C = X^2$$
$$Y^2 + C = Z^2,$$

essendo, secondo la terminologia di Pacioli (ripresa poi da Galigai, Tartaglia, Cardano), $C$ il numero congruente e $Y^2$ il quadrato congruo.

Pacioli insisteva sul problema per una decina di pagine, e in particolare: *a)* dava le risolventi $C = 4ab(b-a)(b+a)$, $Y = b^2+a^2$, essendo $(a,b)$ due numeri interi; *b)* indicava che $C$ era sempre il numero dato e doveva essere intero; il quadrato congruo poteva risultare invece frazionario; *c)* presentava una tabella di coppie $(C, Y^2)$ intere, consigliando di cercarvi il $C$ dato o un suo multiplo quadratico $CN^2$; in questo secondo caso si sarebbe avuta la soluzione frazionaria $y^2 = Y^2/N^2$. Faceva gli esempi per $C = 6$, $C = 30$; *d)* prospettava l'insufficienza di tabelle del genere per molti valori di $C$ e la necessità di ricorrere allora a «regole straordinarie» per giungere alla soluzione. Dava una traccia di tali regole per $C = 7$, ottenendo $y^2 = 7 + 12769/14400$. Dava solo le soluzioni $y^2 = 5 + 97/144$ per $C = 5$, e $y^2 = 13 + 164568241/375584400$ per $C = 13$.

Diventava così chiaro il significato della sua frase che, nonostante tutti gli sforzi di Leonardo Pisano, le soluzioni del problema del congruo «pur si convien a tastoni ridurse in cercarle».

Aggiungiamo che nonostante tutti gli sforzi di Leonardo, Antonio de' Mazzinghi, Giovanni di Bartolo, Maestro Benedetto in periodo medioevale e di Genocchi, Woepcke, Collins, Lucas e del calcolatore in periodo moderno, il pur limitato campo per $C$ minore di 100 non è ancora completamente coperto.

Summa de Arithmetica geo metria. Proportioni: et proportionalita:
Novamente impressa In Toscolano su la riua dil Benacense et vnico carpionista Laco: Amenissimo Sito: de li antique, e evidenti ruine di la nobil cita Benaco ditta illustrato: Cum numerosita de Impatorii epithaphij di antique e perfette littere sculpiti dotato: e cu finissimi e mirabil colone marmorei: inumeri fragmenti di alabastro porphidi e serpentini. Cose certo lettor mio diletto oculata fide miratu digne sotterra se ritrovano.

Continentia de tutta lopera:

De numeri e misure in tutti modi occurrenti.
Proportioni e pportionalita a notitia del 5° de Euclide: e de tutti li altri soi libri.
Chiaui: ouero euidētie numero. 13. per le quantita continue pportionali del 6° e 7° de Euclide extratte.
Tutte le parti de lalgorismo: cioe relevare partire: multiplicare: sūmare: e sotrare: con tutte sue pue in sani e rotti e radici e progressioni.
De la regola mercantesca ditta del. 3. e soi fondamēti cō casi exēplari p c° m° ʒ. G. guadagni: perdite: transportationi: e inuestite.
Partir: multiplicar: sūmar: e sotrar de le pportioni: e de tutte sorti radici.
De le tre regole del Catayn ditta positione: e sua origine.
Euidentie generali: ouer conclusioni numero. 66. absoluere ogni caso che per regole ordinarie non si podesse.
Tutte sorte binomii e recisi: e altre linee irrationali del decimo de Euclide.
Tutte regole de Algebra ditte de la cosa e lor fabriche e fondamenti.
Cōpagnie in tutti modi: e lor partire.
Socide de bestiami: e lor partire.
Fitti. pesciōni: cottimi: livelli: logagioni: e godimenti.
Baratti in tutti modi semplici: compositi: e col tempo.
Cambi real: secchi: fittitii: e diminuti: ouer communi. (termini
Meriti semplici e a capo danno: e altri
Resti: saldi: sconti: de tēpo e denari: e de recare a vn di piu partite.
Ori: argenti: e lloro affinare: e carattare.
Molti casi e ragioni straordinarie: varie e diuerse a tutte occurrētie: cōmo nella sequente tauola appare ordinatamente de tutte.
Ordine a saper tener ogni cōto: e scripture: e del quaderno in vinegia.
Tariffa de tutte vsançe e costumi mercanteschi in tutto il mondo.
Pratica e theorica de geometria: e de li cinqs corpi regulari: e altri depēdenti
E molte altre cose de grandissimi piaceri: e frutto: commo diffusamente per la sequente tauola appare.

Frontespizio dell'edizione di Toscolano della *Summa* di Luca Pacioli, con il contenuto dell'opera. Già vi è evidente l'estrosità dell'autore, che nel corso della trattazione matematica inserisca commenti di varia natura, termini e frasi latine. Ne consegue un'opera *sui generis*, animata da un pittoresco disordine che, forse troppo severamente, rimproveravano al suo autore Tartaglia e Cossali.

Per tre secoli nessuno diede particolare importanza ai passi della *Summa* citati; gli stessi termini congruo-congruente scomparvero dal lessico matematico.

Fu il parmense d'adozione, ingegnere, matematico e storico delle matematiche don Pietro Cossali a rendersi conto dell'importanza degli argomenti trattati e a mettersi alla ricerca del manoscritto perduto di Leonardo Pisano sui numeri quadrati.

Risultate vane le sue ricerche, tentò arditamente la ricostruzione dell'opera, sulla scorta dei dati di Pacioli, in una cinquantina di pagine della sua opera *Origine e trasporto in Italia dell'algebra. Storia critica di nuove disquisizioni analitiche e metafisiche arricchita* edita a Parma nel 1793. Diciamo subito che tale ricostruzione può essere considerata oggi alla stregua di curiosità storica, ivi incluso il tentativo di arrivare in base alle teorie di Eulero-Lagrange alla soluzione «diretta» del problema congruo-congruente, cioè senza dover procedere per tentativi.

Ciò però che colpisce il Cossali è il suo intuito, l'inquadramento che sa dare alla figura e al carattere di Leonardo Pisano, la certezza acquisita in base alla lettura del *Liber abaci* che egli fosse stato un grande della matematica in senso assoluto.

Per tale certezza accusava, e non a torto, di trascuratezza Montucla, il massimo storico della matematica dei suoi tempi: «... non ha poi una parola del libro di Leonardo sui numeri quadrati... con che gli diminuì un ramo di serto di gloria dovutogli, richiedendo verità, che si riconosca primo maestro di queste regioni della determinata e dell'indeterminata analisi ugualmente».

Montucla, si deve ricordare, distribuiva il merito di aver introdotto l'algebra in

El p.° numero congruo e. 25. che receue e dona. 24
2.° 100. che receue e dona. 96.
3.° 169. che receue e dona. 120.
4.° 225. che receue e dona. 216.
5.° 298. che receue e dona. 240.
6.° 400. che receue e dona. 384.
7.° 625. che receue e dona. 336.
e cosi receue e dona. 600.
8.° 676. che receue e dona. 480.
9.° 841. che receue e dona. 840.
10.° 900. che receue e dona. 864.
11.° 1156. che re. e do. 960.
12.° 1225. che. r. e. d. 1176.
13.° 1512. che. r. e. d. 1080.
14.° 1600. che. r. e. d. 1536.
15.° 1681. che. r. e. d. 1720.
16.° 2025. che. r. e. d. 1944.
17.° 2500. che. r. e. d. 1344.
e cosi receue dona. 2400.
18.° 2601. che. r. e. d. 2160.
19.° 2704. che. r. e. d. 1920.
20.° 2809. che. r. e. d. 2520.
21.° 3025. che. r. e. d. 2904.
22.° 3364. che. r. e. d. 3360.
23.° 3600. che. r. e. d. 3456.
24.° 3721. che. r. e. d. 1320.
25.° 4225. ch. r. e. d. piu n'. 2016
E 3000. E 3696. E 4056.
26.° 4624. che. r. e. d. 3840.
27.° 4900. che. r. e. d. 4704.
28.° 5329. che. r. e. d. 5280.
29.° 5476. che. r. e. d. 3360.
30.° 5625. che. r. e. d. 3024.
e cosi receue e dona. 5400.
31.° 6084. che. r. e. d. 4320.
32.° 6400. che. r. e. d. 6144.
33.° 6724. che. r. e. d. 2880.
34.° 7225. ch. r. e. d. piu n'. 2184
E 5544. E 6000. E 6936
35.° 7569. che. r. e. d. 7560.
36.° 7921. che. r. e. d. 6240.
37.° 8100. che. r. e. d. 7776.
38.° 8281. che. r. e. d. 5880.
39.° 9025. che. r. e. d. 8664.
40.° 9409. che. r. e. d. 9360.
41.° 10000. ch. r. e. 5376 | 9600.
42.° 11025. che. r. e. d. 10584.
43.° 12100. che. r. e. d. 11616.

**Tabella di coppie congruo-congruente riportata a margine del foglio 46v della seconda edizione della *Summa* di Luca Pacioli. Consta di 43 congrui e 53 congruenti. Genocchi ha fatto notare che il tredicesimo congruo è 1521. Manca anche la coppia (1389, 840).**

Europa tra Leonardo Pisano, Paolo dell'Abaco e Prosdocimo de Beldomandis, vissuti nell'intervallo di oltre due secoli.

Il capolavoro di Cossali riguarda comunque la sua difesa *a priori* di Leonardo contro chi (era un presagio) avesse osato accusarlo di aver preso da Diofanto che nel II secolo ad Alessandria aveva composto l'*Aritmetica*, l'opera unica di analisi indeterminata dalla quale prenderà avvio con Fermat la matematica moderna.

Lo spunto glielo diede Xilander, primo traduttore dal greco nel 1575 dell'*Aritmetica*, il quale, avendo appreso «che un certo Leonardo Pisano aveva scritto un libro sui numeri quadrati», aveva commentato: «non dubito quin ex nostro transtulerit Diophanto».

Sistemato in breve l'incauto Xilander, Cossali veniva però assalito dal dubbio che Leonardo Pisano avesse conosciuto l'opera di Abul Wafa, matematico persiano del quale sapeva che aveva scritto nel decimo secolo un'opera di commento e una di dimostrazioni (e a dimostrazioni geometriche di Leonardo aveva accennato Pacioli) delle proposizioni di Diofanto.

Espressa l'improbabilità che una tale opera, poco nota agli stessi arabi, fosse invece conosciuta dal matematico pisano, egli passava comunque a un confronto diretto tra le soluzioni date da Diofanto e da Leonardo ad alcuni problemi comuni, per concludere: «Come persuaderci che Leonardo versando l'opera di Diofanto non avesse appresi gli artifizi di Diofanto... Come indurci a credere che tenendo sotto de' suoi occhi la riferita question di Diofanto, in faccia della illimitata ampiezza di esso limitato si fosse e nella proposta, e nello scioglimento? Io non so dunque piegar la mente a pensare, che Leonardo abbia veduta l'opera di Diofanto, e da lui abbia tolto.»

Il binomio Diofanto-Leonardo comparirà spesso in seguito presso storici e studiosi alla ricerca delle «fonti» di Leonardo Pisano; ma sarà sempre l'evidenza della diversità dei procedimenti seguiti dai due matematici per risolvere problemi affini e il raffronto tra «l'illimitata ampiezza» dei procedimenti del matematico greco e quella, geniale fin che si vuole ma limitata nelle sue impostazioni, del matematico toscano, a dare la conferma della originalità e piena indipendenza di Leonardo Pisano da Diofanto.

Se il ritrovare il *Libro sui numeri quadrati* sarebbe stato degno premio per Cossali, non fu certo a danno della scienza se tale premio toccò poi al principe Baldassarre Boncompagni, dottissimo e infaticabile ricercatore e mecenate. Nel Codice E. 75 P. sup. della Biblioteca Ambrosiana di Milano egli trovò ai primi del 1853 non solo l'originale latino del *Liber quadratorum*, ma anche quelli del *Flos* e della *Lettera a Maestro Teodoro*, scritti dei quali neppure si sospettava l'esistenza. Li pubblicò l'anno successivo e fece seguire tra il 1857 ed il 1862 l'edizione completa delle opere di Leonardo Pisano.

Apparve allora in piena evidenza l'originalità del matematico pisano e il livello del suo apporto alla matematica, ben superiore a quanto non apparisse dalla *Summa*, ben al di là di quanto lo stesso Cossali aveva immaginato. «Ho io dovuto, con fatica certamente maggiore di quella che l'originale cagionato avrebbemi, raccoglier le dottrine di Leonardo dal ceneraccio, a dir con Annibal Caro, di Fra Luca», aveva detto il padre teatino. Ma su questo s'ingannava, in quanto una schiera di quotati studiosi, lo stesso Boncompagni, Genocchi, Woepcke, Terquem, Lebesgue, dovettero impegnarsi a lungo e a fondo per venire a capo degli *Opuscoli* (così li chiamò Boncompagni) di contenuto ben più complesso che non le due opere maggiori.

Il mondo scientifico ne rimase sorpreso e ammirato: «Non si immaginava - scriveva Terquem - che un geometra del XIII secolo avesse superato di tanto Diofanto e gli Arabi per non esser superato che nel XVII secolo da Fermat.»

Nelle prefazioni agli *Opuscoli* inoltre Leonardo Pisano spiegava le vicende che lo avevano portato a cimentarsi nel campo dell'analisi indeterminata.

Leonardo Pisano nella sua gioventù aveva visto la sua città espandersi oltr'Arno, l'aveva vista cingersi di mura, aveva visto completare il Duomo con le porte bronzee di Bonanno, aveva visto sorgere inclinata la famosa torre, aveva visto nascere il battistero di Diotisalvi. Mentre attorno al 1223 il Campo dei Miracoli era in via di completamento e a Leonardo la sua città, forte dell'appoggio imperiale, appariva al culmine della sua potenza, Federico II passò per Pisa e nel palazzo imperiale il matematico pisano gli fu presentato da Maestro Domenico.

Davanti all'imperatore si intrattenne in discussioni matematiche con Maestro Giovanni da Palermo, un altro filosofo di corte. Questi propose anche alcuni problemi a Leonardo; uno di essi richiedeva la soluzione di un'equazione di terzo grado, un altro era sul congruo: «Trovare un numero quadrato tale che, sia aggiungendogli che togliendogli 5, dia sempre un quadrato».

Tali discussioni e sfide erano nelle abitudini dei tempi e nei gusti di Federico II; d'altra parte Leonardo si era abituato a sostenerne sin dalla giovinezza. Leonardo riferì a Maestro Giovanni (in breve lasso di tempo, dobbiamo ritenere, in quanto non è difficile trovarla) la soluzione del problema sul congruo: il quadrato cercato era $11+2/3+1/144$, ossia $(41/12)^2$.

Per l'equazione di terzo grado i tempi di risposta furono certo più lunghi, in quanto per dimostrarne l'insolubilità a mezzo di razionali o di irrazionali euclidei dovette studiare a fondo e passare in termini aritmetici l'arduo e lungo X Libro degli *Elementi* di Euclide. Nella lettera con la quale trasmetteva a Federico II i risultati di questi studi Leonardo gli comunicava anche di aver ripreso il «problema del congruo» e di averne messo a punto la teoria in un *Libretto dei quadrati*, a lui dedicato.

Dai regesti di Federico II risulta che egli passò per Pisa solo nel luglio 1226. D'altra parte il *Liber quadratorum* è datato 1225 e quindi, tenendo conto di questa data e dei tempi di esecuzione, si può ritenere che l'incontro sia avvenuto attorno al 1223.

Umberto Forti ritiene che, data la prontezza con la quale Leonardo Pisano seppe trovare la soluzione approssimata dell'equazione di terzo grado, il «torneo» di Pisa fosse «alquanto truccato». Leonardo non dice però di aver dovuto dare risposta immediata ad alcuno dei problemi.

Si deve piuttosto notare che a conclusione della *Pratica geometrie*, qualche anno prima, Leonardo aveva inserito un problema di analisi indeterminata che con la geometria proprio nulla aveva a che vedere. Si trattava di risolvere la $y^2+5=z^2$. Non si può parlare di aggiunta posteriore di un copista, in quanto il latino e la tecnica risolutiva (la stessa con la quale nel *Liber quadratorum* tratterà terne pitagoriche e congruo-congruenti) sono inconfondibilmente suoi.

Si deve allora ritenere che, visto questo problema, Maestro Giovanni abbia voluto metterlo alla prova, imponendo un secondo e impegnativo vincolo, che fosse cioè anche $y^2-5=x^2$.

Questo problema, si è detto prima, non è comunque di difficile soluzione. Basta infatti tabulare le differenze tra i quadrati dei numeri dispari minori di 50 per riscontrarvi le soluzioni relative a sei congrui, e tra esse quella del problema proposto (*si veda la tabella a pagina 102*). Si ha infatti:

$$41^2-720=31^2 \quad (41/12)^2-5=(31/12)^2$$
$$41^2+720=49^2 \quad (41/12)^2+5=(49/12)^2.$$

Passando al *Liber quadratorum*, Leonardo Pisano risolve anzitutto il problema delle terne pitagoriche (trovare due quadrati la cui somma sia un quadrato). Applicando un teorema degli *Elementi* arriva a ottenere l'equivalente della nota identità algebrica

$$(b^2-a^2)^2+(2ab)^2=(b^2+a^2)^2.$$

Adotta però anche un procedimento (basato sulla proprietà dei quadrati di esser somma di dispari successivi a partire da 1), che è originale suo, e che riferiamo a mezzo di un suo esempio.

Si fissi un quadrato qualsiasi, per esempio 81; esso è divisibile per 3 ed è quindi somma dei tre dispari 25, 27, 29 disposti attorno a $27=81/3$. Gli altri due quadrati saranno dati da 144, somma dei dispari da 1 a 23 (il dispari che precede 25), e da 225, somma dei dispari da 1 a 29 (il maggiore dei dispari nei quali si è scomposto 81).

Analogamente, se il quadrato fissato è pari. È lo stesso procedimento che Leonardo aveva adottato per il problema indeterminato nella sua *Pratica geometrie*.

**Il primo foglio del Cod. E-75 P. Sup. della Biblioteca Ambrosiana di Milano. Viene presentato al Cardinale Raniero Capocci il *Flos*, nei seguenti termini: «Incipit flos Leonardi bigolli pisani...». Alla quinta riga dal fondo Leonardo Pisano riprende poi la lettera con la quale aveva trasmesso a Federico II le soluzioni dei problemi che gli erano stati proposti al convegno di Pisa.**

Leonardo passa poi a dimostrare che $ab(b-a)(b+a)$, se $(a,b)$ sono entrambi pari o entrambi dispari, è multiplo di 24, e così $4ab\,(b-a)(b+a)$, se $(a,b)$ sono uno pari e l'altro dispari. Prima dimostra che sono multipli di 8 e poi di 3. Questa seconda parte è notevolmente interessante e originale (il procedimento verrà ripreso ed esteso da Eulero nell'*Algebra*).

I numeri $(a,b)$ possono essere dei tre tipi: $3k$, $3k+1$, $3k+2$. Se uno dei due è di tipo $3k$, la tesi è dimostrata. Se sono entrambi di tipo $3k+1$ o $3k+2$, $(b-a)$ è di tipo $3k$ e la tesi è dimostrata. Se sono uno di tipo $3k+1$ e l'altro di tipo $3k+2$, $(a+b)$ è di tipo $3k$, e la tesi è dimostrata.

Leonardo Pisano conclude: «Ho chiamato congruo, un numero così fatto, e che è sempre multiplo di 24». (Seguiremo questa definizione di Leonardo; Pacioli lo chiama invece, come si è visto, congruente.) Passa a questo punto alla soluzione generale del problema del congruo.

La caratteristica fondamentale e decisiva del congruo individuata da Leonardo è che «esso è in due modi diversi somma di dispari successivi». Riprendendo l'e-

**Relazioni sui triangoli numerici *(a sinistra)*. La (1) con $(a, b)$ uno pari e l'altro dispari, primi tra loro, dà tutte e sole le terne primitive $(d, p, i)$; generali sono la (2) a tre parametri e la (3), a due parametri ma con $h \times k =$ quadrato, dovuta a Euclide. Disposti otto triangoli numerici uguali come a destra e indicando con $A$ l'area di ciascun triangolo $(d, p, i)$, è immediata l'equivalenza tra la (1) medioevale e la (2) di Diofanto. Le (3) danno le espressioni parametriche di $(C, X, Y, Z)$.**

| $X^2$ | $Y^2$ | $Z^2$ | C | c | Y | (a, b) |
|---|---|---|---|---|---|---|
| 1 | 25 | 49 | 24 | 6 | 5 | 1,2 |
| 49 | 169 | 289 | 120 | 30 | 13 | 2,3 |
| 49 | 289 | 529 | 240 | 15 | 17 | 1,4 |
| 289 | 625 | 961 | 336 | 21 | 25 | 3,4 |
| 1 | 841 | 1681 | 840 | 210 | 29 | 2,5 |
| 529 | 1369 | 2209 | 840 | 210 | 37 | 1,6 |
| 961 | 1681 | 2401 | 720 | 5 | 41 | 4,5 |

**Calcolando le differenze tra i quadrati dei dispari minori di 50 già si individuano sette terne di quadrati congruenti, cioè che differiscono fra loro per uno stesso congruo $C$. Dividendo poi $C$ per i fattori quadratici che lo costituiscono, si ottengono i congrui $c$ a quadrati congruenti frazionari. In ultima riga si hanno gli elementi per le soluzioni del problema proposto a Leonardo Pisano. Leonardo individuò tra l'altro le seguenti caratteristiche: $Y$ è somma di due quadrati $a^2 + b^2$; $C$ è sempre multiplo di 24; $24 \times k^2 \times Sx^2$, essendo $Sx^2$ somma di quadrati di successivi interi o di successivi dispari a partire da 1, è un numero congruo (tutti i $C$ in tabella soddisfano a questa condizione).**

**Incontro di Federico II con il suo avversario, il Malik al Kamil, durante la crociata che lo portò all'occupazione patteggiata di Gerusalemme, dal Cod. Chici L VIII 286 della Biblioteca Apostolica Vaticana. Lo storico arabo Ibn Wasil riferisce di difficili quesiti filosofici e matematici inviati dall'imperatore ai dotti arabi durante la crociata. Da un documento pubblicato dall'Amari si sa ancora di quesiti inviati ai dotti musulmani tra il 1230 e il 1240 in relazione alla creazione del mondo e all'immortalità dell'anima. Fondatore nel 1224 dell'Università di Napoli, Federico II instaurò nel meridione d'Italia un vero regime culturale d'avanguardia. Egli stesso scrisse il *De arte venandi cum avibus*, opera di impronta decisamente scientifica che ebbe vasta diffusione. Mentalità scientifica ebbe anche il figlio Manfredi, che aggiunse alcune note all'opera del padre e che, a detta di Giamàl ad Din, che fu presso di lui a Barletta, conosceva a memoria gli *Elementi* di Euclide.**

sempio numerico per $(X,Y,Z,C) = (31, 41, 49, 720)$, si ha:

$$31^2 = 1 + 3 + ... + 59 + 61$$
$$41^2 = 1 + 3 + ... + 79 + 81$$
$$49^2 = 1 + 3 + ... + 95 + 97.$$

Ne risulta la doppia origine del congruo 720:

$$720 = 41^2 - 31^2 = 63+65+...+79+81$$
(dieci dispari)
$$720 = 49^2 - 41^2 = 83+85+...+95+97$$
(otto dispari)

e quindi un doppio modo di esprimerlo come prodotto delle quantità di dispari che lo costituiscono per il loro valor medio:

$$720 = 10 \times 72 = 8 \times 90.$$

Saranno questi quattro elementi, le due quantità di dispari (10 e 8 nell'esempio) e i valori medi attorno ai quali essi sono distribuiti (72 e 80 nell'esempio) i veri capisaldi operativi della trattazione di Leonardo Pisano; in funzione di essi egli esprimerà i lati dei quadrati congruenti, come passo conclusivo della sua ricerca.

Collegando poi tali capisaldi ai parametri $(a,b)$, si potranno esprimere, in funzione di essi, quadrati congruenti e loro lati.

I suoi successori medioevali utilizzeranno solo le espressioni di $(C,Y)$ in funzione di $(a,b)$

$$C = 4ab(b-a)(b+a) \quad Y = b^2+a^2.$$

Questa in sintesi, e riferita a un esempio numerico (non ripreso dalla sua opera), la *teorica dei congrui*, che Leonardo Pisano sviluppa attraverso complesse e rigorose dimostrazioni geometriche e con distinzioni di casi e sottocasi (per distinguere le due situazioni per noi traducibili come $X > 0$, $X < 0$).

Nulla di simile si trova nelle trattazioni di Diofanto e degli arabi per giungere alle risolventi del problema.

Gli storici della matematica hanno fortunatamente continuato a indicare il nostro come Leonardo Pisano, con lo stesso nome cioè con cui lo ricordano tutti gli scritti medioevali e rinascimentali. Ma per i matematici è Leonardo Fibonacci, e più che per la sua opera è forse noto per la serie ricorrente (nella quale ogni termine è somma dei due che lo precedono):

$$(1), 1, 2, 3, 5, 8, 13, 21, 34, ...,$$

riportata da Leonardo a margine del testo del noto «problema dei conigli» del *Liber abaci*, un gioco più che un vero problema. I matematici moderni ne hanno individuato importanti proprietà, l'hanno battezzata «serie di Fibonacci» e chiamano «numeri di Fibonacci» i suoi elementi; a ogni modo Fibonacci è il cognome che il matematico pisano non ha mai avuto in vita.

Principale responsabile dell'iniziativa fu certamente Guillaume Libri, che in una nota a pagina 19 del secondo volume della sua *Histoire des sciences mathématiques en Italie* (1838) così la giustificava: «Fibonacci è una contrazione de filius Bonacci, contrazione della quale si trovano numerosi esempi nella formazione dei cognomi delle famiglie toscane.»

In effetti nelle cronache pisane dal XII al XIV secolo si trovano a Pisa dei «de Bonaccis», dei «de Bonagis», ma nessuna traccia dei Fibonacci. Se ne fa però menzione nei *Ricordi di ser Perizolo* del 1510, come riportato in una memoria del 1867 dal Bonaini: «Lionardo Fibonacci fue nostro concive e vivette gli anni 1203, vidde tutto il mondo; tornoe a Pisa e rechò i numeri arabichi e l'aritmetica e ne compose un libro...».

Il cognome Fibonacci gli risparmiò comunque quello di Bigollo, ancor meno riverente seppur più legale del primo, in quanto egli stesso così presenta il *Flos*: «*Incipit flos Leonardi bigolli pisani...*».

Per iniziativa ancora di Libri, furono poi tirati in causa e messi sotto accusa i concittadini di Leonardo Pisano. Il manoscritto del *Flos* era allora sepolto negli scaffali dell'Ambrosiana, ma rifacendosi all'*Elogio di Leonardo Pisano* del Guglielmini del 1813, e a un manoscritto della *Pratica geometrie* da lui trovato alla Biblioteca Reale di Parigi «... Composita a Leonardo Bigollosio fillio Bonaccij», Libri commentava: «In premio degli immensi servizi che aveva resi alla scienza gli si diede il soprannome [sobriquet] di Bigollone». Ricordando poi l'epiteto di «Messer Millione» dato dai veneziani a Marco Polo: «... Si è visto chiamare dai Pisani fannullone il padre dell'algebra moderna...».

Una ventina di anni dopo (ormai Leonardo Pisano era tornato alla ribalta mondiale) sugli «Annali di scienze matematiche» Terquem rincarava la dose: «Gli uomini superiori sono spesso considerati dei grulli dagli uomini inferiori. È così che i negozianti di Pisa, compatrioti di Leonardo, gli hanno dato il soprannome di Bighelone».

A far prosciogliere i pisani dall'accusa provvide il Bonaini, rinvenendo nell'Archivio fiorentino di Stato un documento del 1241 con il quale, essendo «Potestà del Comune Ugone Rossi da Parma» (le «provincie parmensi» davano così inizio a una specie di gemellaggio con Pisa in difesa di Leonardo, che Cossali e Genocchi avrebbero consolidato), si assegnava uno stipendio annuo di 20 mila lire di denari al «discreto e sapiente Maestro Leonardo Bigollo» per i servizi resi al Comune come consulente.

Ne conseguiva che Bigollo non era un soprannome e che i pisani avevano avuto orgogliosa stima della capacità del loro concittadino, onorandolo persino del titolo di «Magister», del quale si fregiavano i dotti della corte imperiale.

Ignorando il documento, qualche decennio dopo Moritz Cantor adombrava romantiche dipartite di Leonardo, diventato filosofo di corte, in Terrasanta, oppure, se rimasto in patria, nella guerra civile del 1228-1229. In realtà sappiamo che Federico nel 1229, tornato dalla Palestina, dava inizio con i suoi crociati, saraceni e scomunicati, alla marcia verso nord, sappiamo che i suoi «philosophi vagantes» lo seguirono, ma tra essi non era certo il sessantenne Leonardo, che d'altra parte di viaggi ne aveva fatti a sufficienza.

Nel 1228, mentre Federico partiva per la crociata, egli procedeva invece alla ristesura del *Liber abaci;* a indurvelo erano state ancora una volta le insistenze di un filosofo di corte, Michel Scotto, il mago-astrologo traduttore di Aristotele, che Dante gratificherà dell'inferno assieme a Guido Bonatti e al loro dotto imperatore.

L'ultima informazione del 1241 ci mostra comunque il «matematico fannullone» ormai settantenne ancora al lavoro come consulente tecnico del suo Comune. Poi il silenzio definitivo cala su di lui.

Il 1241 è anche l'anno dell'ultima vittoria di Federico II e delle galere pisane, e segna l'inizio della fine della repubblica marinara toscana e della casa sveva. In quell'anno le avanguardie mongole di Batu lo Splendido erano già oltre l'Oder e sull'Adriatico; non fu l'inizio della fine della civiltà europea solo per intervento diretto del Padreterno.

Come risulta da due manoscritti di Maestro Benedetto da Firenze, nel periodo medioevale furono individuati nel campo dei congrui minori di 100 molti quadrati congruenti; l'impresa era particolarmente ardua, sia perché intervenivano spesso grandi numeri (fu determinato, per esempio, in

4079+2856446223180462864l/
89777960534325799600

il quadrato congruente a 37), sia perché, fissato un numero, non si disponeva di alcun criterio per decidere se fosse congruo, e quindi in caso negativo ci si sarebbe inutilmente esposti al faticoso lavoro di ricerca del suo quadrato congruente.

L'unico criterio di esclusione noto sino a metà del secolo scorso si doveva ancora una volta a Leonardo Pisano: «Nessun numero quadrato può essere un congruo».

Siccome poi, come si è visto, l'espressione del congruo $4ab\,(b^2-a^2)$ dà anche il quadruplo dell'area di un triangolo numerico, la proposizione equivaleva ad affermare che l'area di un triangolo numerico non può essere un quadrato; era questa la proposizione della quale Fermat aveva fornito una nota dimostrazione, a mezzo di una lunga e complessa trattazione «per discesa infinita».

D'altra parte Leonardo Pisano, dopo aver dimostrato a mezzo di una lunga e complessa trattazione geometrica che la proposizione $(b+a):(b-a) = b:a$ non può esser soddisfatta per valori razionali di $(a, b)$ (risulta infatti $b/a = 1 \pm \sqrt{2}$), concludeva in questi semplici termini: «Da questo si mostrerà poi che nessun numero quadrato può essere un congruo, in quanto se fosse possibile lo sarebbe anche la proposizione: la somma di due numeri sta alla loro differenza come il maggiore di essi sta al minore».

Per Leonardo quindi l'impossibilità di $ab(b-a)(b+a)$ = quadrato conseguiva dalla impossibilità da lui dimostrata in precedenza di $b(b-a) = a(b+a)$.

Gli studiosi del secolo scorso conclusero, ed era logico, che questa non poteva considerarsi una piena dimostrazione.

Prese allora la difesa di Leonardo il matematico piacentino Angelo Genocchi, preoccupato anzitutto di allontanare dal «*Magister subtilitatum*» il sospetto di non essersi accorto dell'incompletezza della sua dimostrazione. A sostegno della sua tesi Genocchi segnalava che Leonardo con quel «si mostrerà» iniziale *(ostendetur)* aveva certo inteso indicare che la sua era solo una traccia di dimostrazione; anche se, si deve osservare, per noi moderni la traccia è labile.

Scriveva Genocchi: «Dotato di mirabile perspicacia, ammaestrato da studi profondi fatti sopra la geometria d'Euclide... avvezzo a indagare a fondo ogni questione e a non accettar senza prove per vera alcuna proposizione o regola, non poteva non accorgersi che la dimostrazione dipendeva da una proposizione ausiliaria e che questa voleva esser dimostrata. Non è quindi probabile che abbia creduto equivalente a una piena dimostrazione le parole sue dianzi recate...». Si tratta comunque di un giudizio sul matematico pisano, importante per la particolare competenza di chi lo ha emesso in fatto di analisi indeterminata, per la profonda conoscenza degli *Opuscoli* dimostrata nelle sue *Note analitiche*, per le estensioni da lui date ai risultati di Leonardo.

«Reca meraviglia ch'egli lasci tal proposizione ausiliaria senza una parola per provarla ed illustrarla», commentava ancora Genocchi. E la meraviglia resta, pur tenendo conto del fatto che egli non poteva dare all'argomento quell'importanza che gli diedero i posteri.

ORDO · POPVLVSQ · PISANVS
QVO · VETERI · CONTEMPTVS · CALVMNIA · OBTRITA
MAIORVM · VOLVNTAS · STVDIVMQVE
IN · LEONARDVM · FIBONACCIVM
NOTIORA · TESTATIORAQ · ESSENT
AMPLISSIMI · PISANAE · REIPVB · DE · EODEM · DECRETI · EXEMPLAR
QVOD · VNVM · SVPEREST · TANTO · VIRO · AEQVALE · MONVMENTVM
IN · PATRIO · TABVLARIO · AFFIGI · IVSSIT · A · MDCCCLXV ·

CONSIDERANTES · NOSTRE · CIVITATIS · ET · CIVIVM
HONOREM · ATQVE · PROFECTVM · QVI · EIS · TAM · PER
DOCTRINAM · QVAM · PER · SEDVLA · OBSEQVIA · DISCRETI
ET · SAPIENTIS · VIRI · MAGISTRI · LEONARDI · BIGOLLI
IN · ABBACANDIS · ESTIMATIONIBVS · ET · RATIONIBVS
CIVITATIS · EIVSQVE · OFFICIALIVM · ET · ALIIS · QVOTIES
EXPEDIT · CONFERVNTVR · VT · EIDEM · LEONARDO
MERITO · DILECTIONIS · ET · GRATIE · ATQ · SCIENTIE · SVE
PREROGATIVA · IN · RECOMPENSATIONE · LABORIS · SVI
QVEM · SVBSTINET · IN · AVDIENDIS · ET · CONSOLIDANDIS
ESTIMATIONIBVS · ET · RATIONIBVS · SVPRADICTIS · A
COMVNI · ET · CAMERARIIS · PVBLICIS · DE · COMVNI
ET · PRO · COMVNI · MERCEDE · SIVE · SALARIO · SVO
ANNIS · SINGVLIS · LIBRE · XX · DENARIORVM · ET
AMISCERIA · CONSVETA · DARI · DEBEANT · IPSEQVE
PISANO · COMVNI · ET · EIVS · OFFICIALIBVS · IN
ABBACATIONE · DE · CETERO · MORE · SOLITO
SERVAT · PRESENTI · CONSTITVTIONE · FIRMAMVS ·

**Dopo aver eretto il monumento a Leonardo, i pisani collocarono nell'atrio del Grande Archivio la «memoria unica sincrona» scoperta da Francesco Bonaini incisa nel marmo e preceduta da un'iscrizione di Michele Ferrucci. «Calumnia obtrita», la lapide fu scoperta il 6 giugno 1867 in occasione «del meraviglioso spettacolo della triennale *Luminaria*»: così in un opuscolo commemorativo del tempo.**

Con una lunga e complessa trattazione geometrica Leonardo presentava nel *Libro dei quadrati* le relazioni

$$(a^2+b^2)(c^2+d^2) = (ac \mp bd)^2+(bc \pm ad)^2$$

mostrando anche come si possano avere una terza e una quarta scomposizione in somma di due quadrati, quando rispettivamente uno o entrambi i fattori al primo membro siano dei quadrati.

Commentava Terquem nel 1856: «Leonardo dimostra perfettamente questa proposizione che, secondo l'osservazione di Woepcke, gli appartiene. Diofanto può aver conosciuto questa proprietà, ma non l'ha enunciata, e la dimostrazione, soprattutto per via grafica, non è facile. Il nome di Fibonacci deve restar legato a questo teorema.» E Loria tornava a insistere nel 1929: «... in memoria di chi per primo lo scoperse, meriterebbe il nome di Teorema di Fibonacci».

Dello stesso parere non era Ver Eecke, che nel 1952 scriveva: «Questa proposizione solleva una questione storica di priorità in quanto enuncia in termini velati le identità attribuite a Lagrange.»

A questo punto, tenendo conto sia del fatto che l'enunciato di Leonardo era chiarissimo, sia che la proposizione ai tempi di Lagrange era ormai diventata di pubblico dominio (era stata ritrovata da Viète e da Bachet de Méziriac oltre un secolo prima ed era stata poi estesa da Eulero), la priorità rivendicata a Leonardo sembrava legittima.

Poteva al più lasciare qualche dubbio il fatto che il matematico pisano avesse verificato ma non ricavato la sua relazione, d'altra parte ben difficilmente deducibile per generalizzazione da casi numerici.

E in effetti nel 1979 Rashdi Rashed ha pubblicato uno studio dal quale risulta che al-Khâzin, matematico arabo del X secolo, prendendo spunto dalla Proposizione III-19 di Diofanto, aveva ricavato la risolvente (aggiungendo e togliendo $2abcd$ allo sviluppo del primo membro) ed era poi passato ad analizzare vari casi e modi di scomposizione di un numero in somma di due quadrati.

Tale quindi l'avvio a una nuova analisi sulla natura del numero, partendo da un problema che solo con Fermat e Gauss avrebbe avuto piena soluzione.

Nel 1860-1861 Woepcke pubblicava la traduzione con commento di due manoscritti arabi del X secolo (uno degli autori è anonimo, l'altro è al Hosein) di contenuto quasi identico. Vi si ricava la formula delle terne pitagoriche per passare poi a dedurne le formule risolventi del problema del congruo; sono quelle già date da Diofanto, ma i due autori ricavano le risolventi partendo direttamente dagli *Elementi*. L'anonimo conclude con una tabella che riporta 34 valori di $(a,b)$, i corrispondenti triangoli primitivi e le relative coppie di congruo-congruente: in definitiva con una tabella del tipo di quella da cui Pacioli consigliava di partire.

Gino Loria fu profondo estimatore di Leonardo Pisano e della sua opera e sostenne, in particolare, che se il *Liber quadratorum* fosse stato conosciuto prima la moderna teoria dei numeri ne avrebbe avuto quella spinta che le impresse poi Fermat; dopo aver esaminato i due scritti arabi e la prima parte del *Liber quadratorum* (terne pitagoriche) egli però così commentava: «Se sembra difficile negare che alle ricerche testè compendiate il Pisano sia stato indotto dall'esempio di Muhammed ibn Hosein, la sua dipendenza da questo si mostra ancor meno dubbia riguardo alla sezione seguente del *Liber quadratorum*, la quale tratta dei numeri congrui.»

Ora, si deve dire che anzitutto è urtante dover pensare al maestro delle «*subtilitates*» che «dipende» proprio da al Hosein, il più modesto tra i due autori arabi e incline a errori anche su questioni elementari, tanto da colpire lo stesso Woepcke per le sue «méprises et inadvertances». E poi, a parte l'improbabilità che Leonardo conoscesse proprio questi due studi (gli unici di parte araba a noi noti sino al 1979 sul problema del congruo), non si riesce a trovare tra le trattazioni dei due arabi e quella di Leonardo Pisano affinità alcuna, né per quanto riguarda le terne pitagoriche, né per quanto riguarda i congrui.

Ci sembra piuttosto degna di segnalazione un'indagine dell'anonimo arabo (che la dichiara propria) sulla natura delle ipotenuse dei triangoli rettangoli, e la sua conclusione che esse sono tutte incluse tra i numeri del tipo $12m+1$ e $12m+5$, pur non essendo tutti i numeri di questi tipi ipotenuse.

In effetti, per esser somma di due quadrati un numero deve includere tra i suoi fattori primi almeno uno del tipo $4n+1$, e inoltre nessuno di quelli del tipo $4n-1$ deve trovarsi elevato a potenza dispari; ma anche per questa conclusione si doveva attendere Fermat.

Il *Liber quadratorum* fu tradotto per la prima volta in lingua moderna, in francese con note, nel 1952 dal belga Paul Ver Eecke; seguì nel 1974 quella (non integrale) in inglese con note dell'americano Edward Grant; nel 1980 è poi apparsa a cura dello scrivente l'edizione commentata italiana (la versione è quella in volgare del XV secolo del Codice Palatino 577 ora alla Biblioteca Nazionale di Firenze).

Ver Eecke racconta nella prefazione di aver ricevuto attorno agli anni venti una copia del *Liber quadratorum* da Ettore Bortolotti, che gliela aveva inviata esprimendogli il desiderio che volesse tradurlo. Sfogliata l'opera, Ver Eecke, impegnato allora nella traduzione delle opere matematiche greche, l'aveva riposta per riprenderla solo all'annuncio della morte dell'amico.

Ricordiamo, in quanto questo può aver almeno in parte condizionato la sua presa di posizione nei confronti del grande pisano, che nel 1952 Ver Eecke era ormai un traduttore-commentatore di fama mondiale, esaltato e premiato per la pubblicazione in francese delle opere dei grandi matematici della Grecia classica (Archimede, Apollonio, Diofanto, Pappo, Proclo, Euclide e altri minori).

Il suo programma d'azione nei confronti di Leonardo si delinea subito nell'introduzione. Premesso di aver trovato il suo latino affetto da «arabismi» (l'unico arabismo cui accennerà in seguito sarà l'espressione «*ex ductu .e. in .z.*», che certo arabismo non è), passa poi ad accennare a due commentatori arabi di Diofanto (Abul Wafa e Kusta ben Luca), le opere dei quali avrebbero potuto influire su Leonardo attraverso scritti o lezioni di matematici arabi a lui contemporanei.

Da questo sfondo si stacca il sensazionale annuncio che Leonardo aveva preso da Diofanto le soluzioni del problema a lui proposto da Maestro Giovanni: «La soluzione in numeri frazionari è data senza procedimento algebrico, partendo dai tre numeri, 31, 41, 49 i cui quadrati sono la progressione aritmetica di ragione 720, numero congruo. L'autore non dà spiegazione sulla scelta di questi tre numeri che non sono né arbitrari né d'intuizione, ma che egli manifestamente prende indirettamente da un problema che si trova in Diofanto.»

In una lunga nota di commento alla Prop. XIV, ripetendo l'accusa, scioglie poi l'enigma sul dove Leonardo Pisano sia andato a prendersi tali soluzioni: «Ce n'est pourtant pas un énigme; car nous avons trouvé qu'il emprunte ces nombres...»; ha trovato la fonte, la Prop. III, 7 dell'*Aritmetica* di Diofanto, via qualche ignoto commentatore arabo, essendone sconosciuti nel XIII secolo sia l'originale greco che sue traduzioni latine.

In realtà Leonardo Pisano non ha lasciato ai posteri alcun enigma da sciogliere; né era necessario avere in lui la fiducia di un Cossali, per rendersi conto che non poteva certo aver tentato di ingannare Federico e la sua dotta corte, e per rendersi conto che una simile grossolana lacuna non avrebbe potuto sfuggire alla competenza e diligenza di un Woepcke o di un Genocchi.

Nonostante la sicurezza ostentata nei suoi commenti, è da essi evidente che Ver Eecke si è trovato più volte in difficoltà nel suo confronto con il latino e la matematica medioevale; sfuggitagli agli inizi la definizione di numero congruo (ritiene e ripete in varie note, e fa dire a Leonardo Pisano, che qualunque multiplo di 24 è un congruo), non può di conseguenza estrarre dalla terribile e unica Prop. XI sulla teorica dei congrui il procedimento attraverso il quale Leonardo vi insegna a calcolare congruo e suoi congruenti.

Leonardo d'altra parte, avendo già mostrato in quattro esempi numerici come si debba procedere, lascia al lettore i passaggi del quinto, consistenti nel ripetere pedissequamente per $(a,b) = (4,5)$ i passaggi da lui effettuati per ricavare da $(a,b) = (1,2)$ i valori $(C,X,Y,Z) = (24,1,5,7)$. E ne sarebbero usciti assieme al congruo 720 gli enigmatici $(X,Y,Z) = (31,41,49)$.

Non si può passar sotto silenzio a questo punto che Grant nella edizione americana del *Liber quadratorum* del 1974

**Angelo Genocchi (Piacenza 1817 - Torino 1889), matematico di fama internazionale, fu dal 1857 titolare della cattedra di analisi e geometria all'Università di Torino. I suoi più notevoli apporti riguardano l'analisi indeterminata, i principi della geometria, lo studio delle serie, il calcolo integrale. Le sue *Note analitiche* del 1855 sono una completa versione algebrica degli *Opuscoli* di Leonardo Pisano, arricchita da commenti, confronti critici, estensioni. L'immagine qui riprodotta è conservata presso la Biblioteca comunale di Piacenza.**

accoglie di fatto la tesi di Ver Eecke, premettendo: «La sua traduzione è stata estremamente utile per interpretare i passaggi difficili e per le sue utili note, che sono state per la maggior parte tradotte in inglese e incluse».

In realtà Grant, esperto di latino e matematica medioevale, ha corretto quasi tutti gli errori della versione francese e eliminato buona parte dei commenti incongruenti. Purtroppo, per fiducia in Ver Eecke o per convinzione, riporta poi integralmente e testualmente alla Prop. XIV la nota di accusa dello studioso belga a Leonardo Pisano, accusa che diventa indiscutibile per il lettore non avendo Grant inclusa la traduzione della pesante ma decisiva Prop. XI.

Si deve infine segnalare che non sembra che in trent'anni opposizione alcuna sia stata avanzata all'interpretazione-accusa di Ver Eecke-Grant.

Ora, una tale opposizione non potrebbe esser sensatamente sollevata da parte di chi non disponesse del non facilmente reperibile testo latino.

Infatti la limpida frase latina «... 144, in quo divide quadratos congruentes eidem 720, quorum primus est 961...», con la quale Leonardo Pisano dice di dividere per 144 i tre quadrati congruenti interi di 720 per ottenerne quelli frazionari di 5, viene deformata in entrambe le versioni in modo da fargli confessare alla Prop. XIV conclusiva di aver scelto, o distanziato di 720 (e quindi di non aver calcolato) dei quadrati. Nessuno quindi senza testo latino potrebbe sensatamente obiettare alcunché a due grandi accusatori e a un grande reo confesso.

# Nei quasar l'immagine del primo universo

*La luce di questi oggetti misteriosi ha impiegato 15 miliardi di anni a giungere sulla Terra. I quasar costituiscono, quindi, una singolare testimonianza di come era l'universo ai suoi primordi*

di Patrick S. Osmer

A diciannove anni di distanza dalla loro scoperta da parte di Maarten Schmidt i quasar costituiscono ancora uno dei grandi enigmi dell'astronomia. Pur se la loro natura resta controversa, non lo è la loro descrizione: i quasar sono oggetti stellari con un grande spostamento verso il rosso. La luce emessa da essi viene fortemente spostata verso l'estremità rossa dello spettro, il che sta a significare che essi si stanno allontanando con una notevole frazione della velocità della luce. Se i quasar sono così lontani come i loro grandi spostamenti verso il rosso parrebbero indicare, essi dovrebbero essere molto più lontani delle galassie ordinarie, i cui spostamenti verso il rosso dimostrano l'espansione generale dell'universo. Un quasar può essere 1000 volte più luminoso di un'intera galassia formata da 100 miliardi di stelle. La luce proveniente dai quasar più lontani ha iniziato il proprio viaggio quando l'universo aveva solo un quarto della sua età attuale e ha impiegato 15 miliardi di anni per arrivare fino a noi.

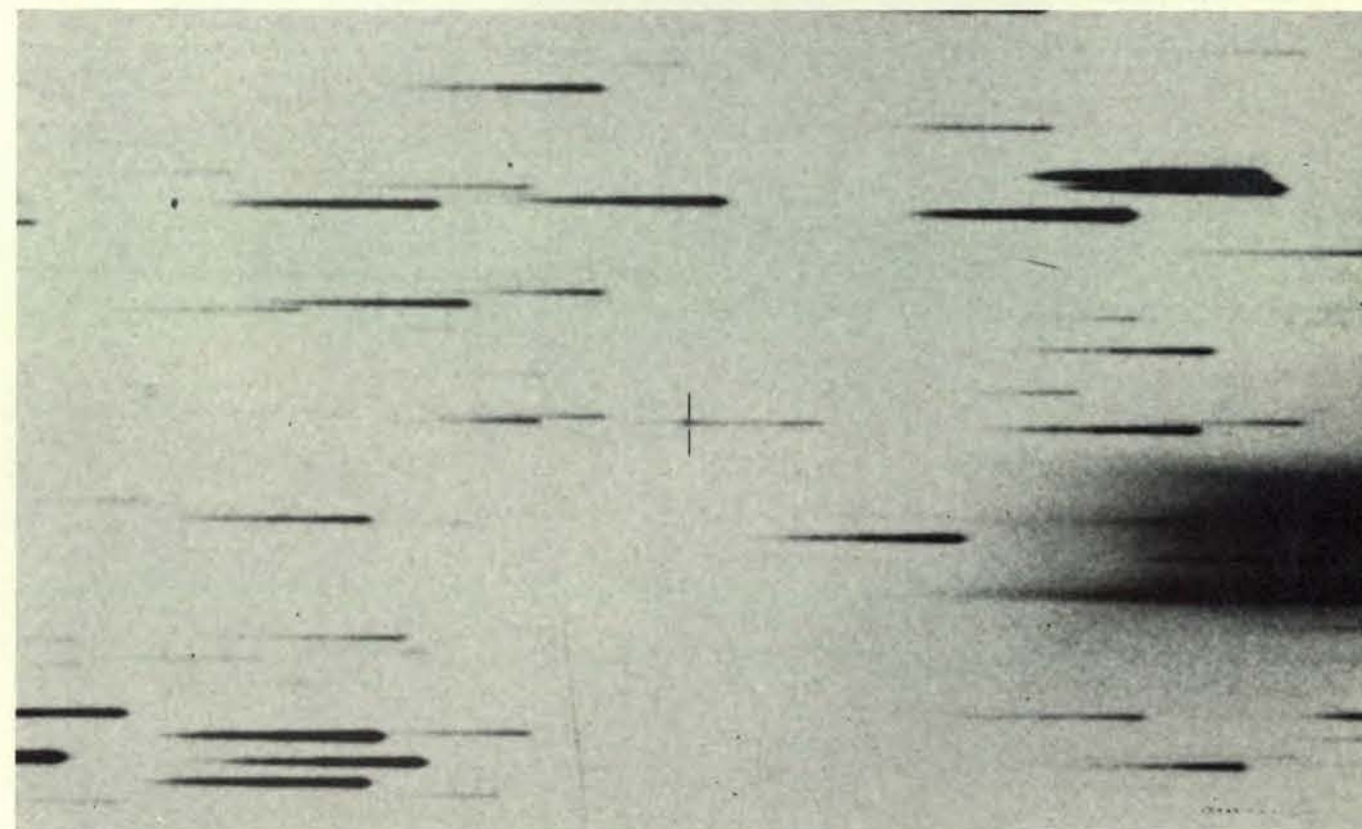
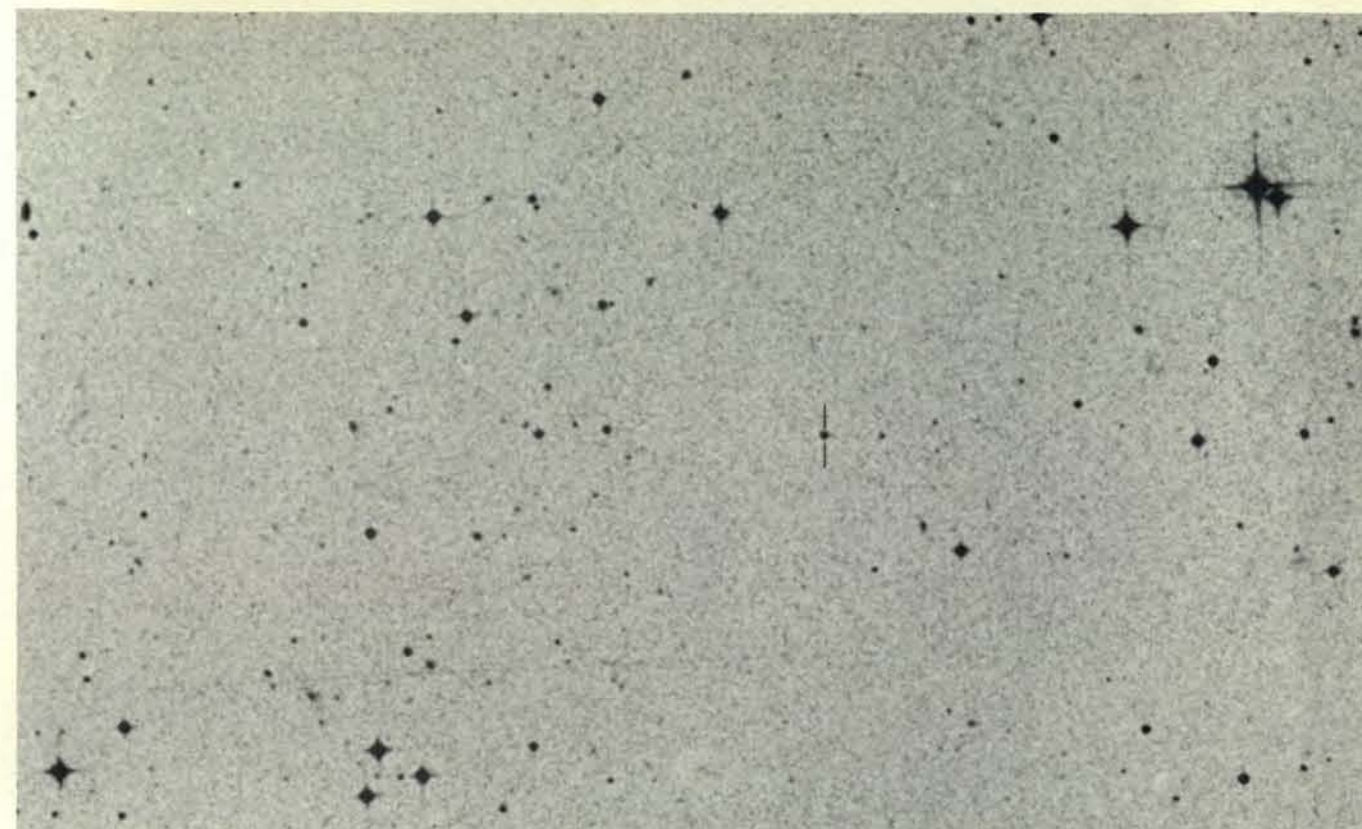
In questo articolo non è mia intenzione discutere la difficile fisica della natura dei quasar; desidero invece concentrarmi su ciò che i quasar possono dirci sul lontano universo e sulle sue condizioni in un'epoca precedente. Adotterò l'ipotesi di lavoro che i quasar siano i luminosissimi nuclei di galassie altrimenti troppo lontane per essere osservate. Attorno a qualche debole quasar relativamente vicino si possono rilevare tracce di una galassia, ma i luminosi quasar lontani si possono distinguere dalle stelle soltanto per le loro proprietà spettrali. La distinzione tra quasar e stelle ha posto gli astronomi davanti a una sfida più impegnativa. Una singola fotografia ad ampio campo del cielo eseguita con un grande telescopio Schmidt (così chiamato dal nome del sistema ottico progettato negli anni venti da Bernhard Schmidt) mostra almeno 200 000 immagini stellari, delle quali solo poche centinaia sono quasar.

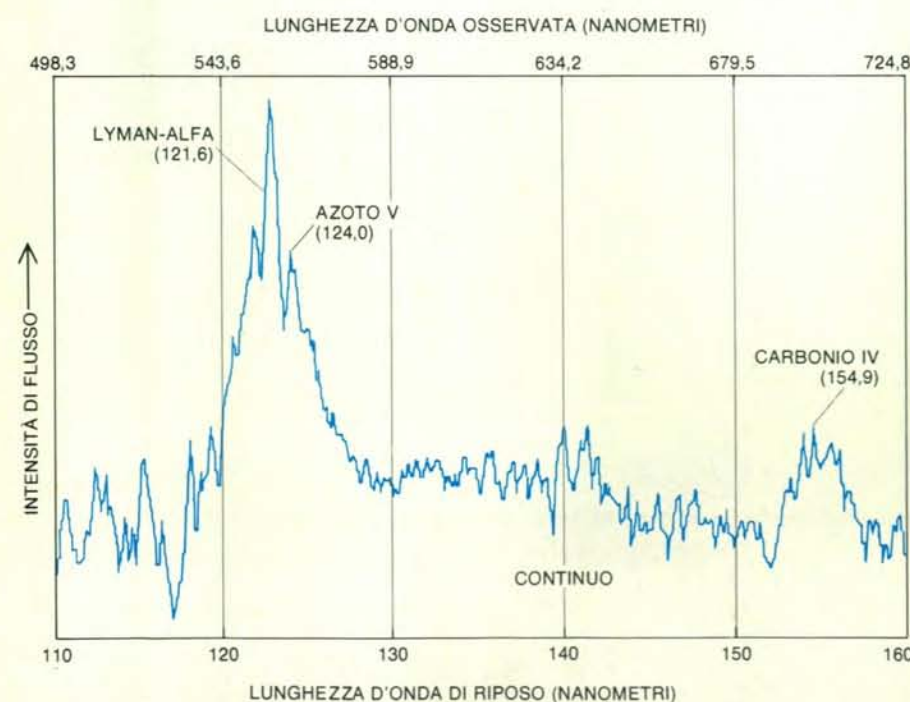
I primi quasar furono scoperti da astronomi ottici che stavano tentando di identificare radiosorgenti celesti. Con il perfezionamento delle tecniche radio, le posizioni assegnate alle radiosorgenti divennero sempre più precise e in molti casi indicavano particolari immagini stellari sulle lastre fotografiche. Dal momento che le stelle normali non potevano essere rivelate come radiosorgenti con le apparecchiature disponibili all'epoca, la coincidenza di una radiosorgente con un oggetto stellare costituì un buon metodo per separare i quasar dalle stelle. Il risultato fu che la maggior parte dei quasar identificati inizialmente erano intense radiosorgenti, quindi radiosorgenti quasi stellari, o quasar.

Si scoprì, però, ben presto che un numero molto maggiore di quasar erano deboli emettitori radio ed erano perciò sfuggiti al rilevamento con radiotelescopi. Allan R. Sandage, lavorando con il telescopio Hale da 200 pollici del Monte Palomar, scoprì che i quasar emettevano molta più radiazione ultravioletta delle stelle ordinarie e potevano pertanto essere individuati confrontando le immagini stellari su lastre fotografiche sensibili all'ultravioletto con le immagini corrispondenti su lastre ordinarie, che sono sensibili principalmente alla luce dell'estremità blu dello spettro visibile. Sandage dimostrò che i quasar che non emettono alle lunghezze d'onda radio sono molto più numerosi di quelli con intense radioemissioni.

Recentemente, per scoprire quasar con grande spostamento verso il rosso, è stato messo a punto un nuovo metodo ottico il quale è complementare al metodo basato sulla loro luminosità ultravioletta. Tale metodo è derivato casualmente da un progetto che io e Malcolm G. Smith avviammo parecchi anni fa quando eseguimmo una ricerca di certe galassie i cui spettri presentano intense righe di emissione anziché righe di assorbimento, che sono più caratteristiche. La ricerca fu eseguita con il telescopio da 60 centimetri Curtis Schmidt dell'Osservatorio interamericano di Cerro Tololo in Cile, congiuntamente a due telescopi a riflessione, di maggiori dimensioni, da 1,5 e da quattro metri di diametro. Smith aveva programmato di fotografare gli spettri di oggetti deboli in un vasto campo disponendo un prisma sottile di nuovo progetto al di sopra dell'apertura del telescopio Schmidt.

Il metodo di disperdere in spettri le immagini normalmente puntiformi delle stelle mediante un prisma obiettivo ha avuto una lunga e caratteristica storia in astronomia stellare. La novità del metodo di Smith fu la combinazione di un prisma a bassa dispersione (per conservare la luce da oggetti deboli) con lastre fotografiche di nuovo tipo con grana fine e alto contrasto. Questa combinazione, insieme alle eccezionali condizioni di osservazione di Cerro Tololo, consentì di registrare spettri di oggetti più deboli di quelli ottenuti con un telescopio di modeste dimensioni come quello Curtis Schmidt. Una volta identificate con questo metodo le galassie, si potevano esaminare dettagliatamente le loro caratteristiche spettrali negli spettri ad alta dispersione forniti dai telescopi più grandi.

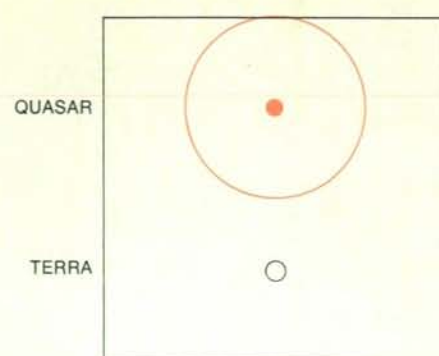
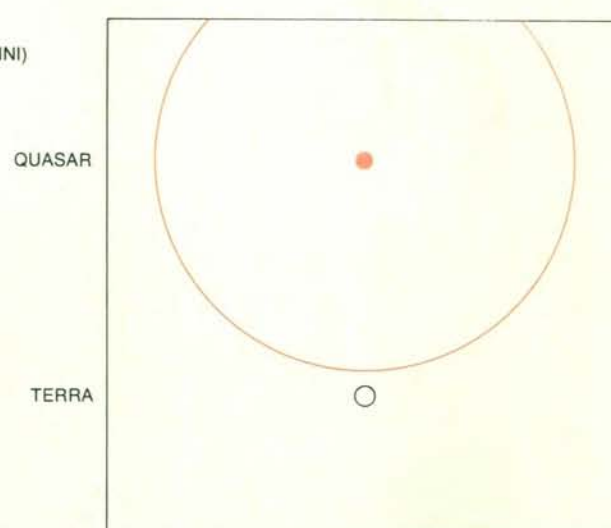
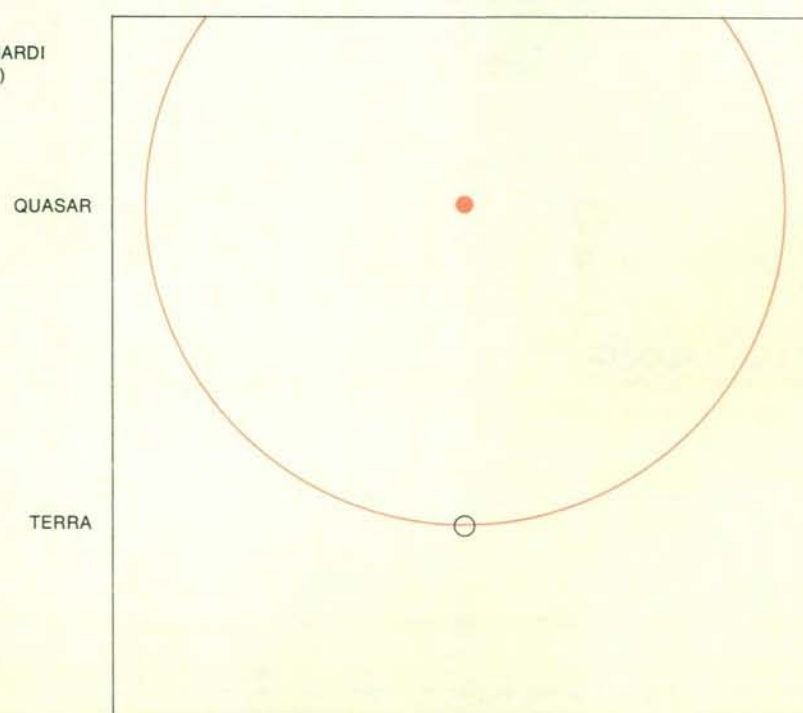
Esaminando le lastre Schmidt ad ampio campo Smith notò che oltre alle galassie con righe di emissione che speravamo di trovare c'erano alcuni oggetti, probabilmente quasar, con righe di emissione in posizioni impreviste nei loro spettri. Quando esaminammo tali oggetti più in dettaglio con telescopi di maggiori dimensioni e con uno spettrometro Vidicon (telecamera intensificata) che avevo adattato allo scopo scoprimmo che si trattava veramente di quasar con grande spostamento verso il rosso. La tecnica di Smith fornì quindi un metodo diretto ed efficace per distinguere i quasar dalle stelle in lastre ad ampio campo e, per distinguere i quasar con grande spostamento verso il rosso, il metodo si rivelò più efficace di quello basato sulla luminosità ultravioletta.

**Il metodo per la ricerca dei quasar è illustrato da queste due fotografie eseguite con il telescopio Curtis Schmidt da 60 centimetri dell'Osservatorio interamericano di Cerro Tololo in Cile. In una singola fotografia diretta del cielo i quasar e le stelle sono indistinguibili perché appaiono entrambi quali punti luminosi come nella stampa in negativo in alto. Quando davanti all'apertura del telescopio Schmidt viene posto un sottile prisma, l'immagine di ogni oggetto viene dispersa in uno spettro, come nella stampa in negativo della stessa porzione di cielo in basso. In quest'ultima fotografia le lunghezze d'onda più lunghe nell'estremità rossa dello spettro sono sulla destra e le lunghezze d'onda più corte nell'estremità violetta dello spettro sono sulla sinistra. Gli spettri delle stelle o sono privi di particolari caratteristiche o presentano bande bianche che indicano l'assorbimento di una parte dell'emissione stellare. Il quasar, indicato da trattini in entrambe le fotografie, presenta un'intensa riga di emissione, la riga Lyman-alfa dell'idrogeno, che è stata spostata dalla sua posizione normale a 121,6 nanometri nella parte dello spettro dell'ultravioletto lontano alla soglia dello spettro visibile a 372 nanometri, che corrisponde a uno spostamento verso il rosso di 2,06. Questo quasar, QO 149-397, è stato uno dei primi scoperti da M. G. Smith con il prisma montato sul telescopio Curtis Schmidt.**

**Gli spettri dei quasar rivelano che gli atomi emittenti sono più altamente ionizzati (privati di più elettroni) degli atomi delle nebulose circostanti le calde e giovani stelle della stessa nostra galassia. Anche le righe di emissione sono fortemente spostate verso l'estremità rossa dello spettro rispetto alla loro lunghezza d'onda di riposo, non spostata verso il rosso. Questo spettro del quasar QO 453-423 è stato eseguito con il sistema di telecamera Vidicon sul telescopio da 1,5 metri di Cerro Tololo. La intensa riga Lyman-alfa ha subito uno spostamento verso il rosso a 445,1 nanometri, nella regione blu dello spettro visibile, a partire dalla sua lunghezza d'onda di riposo di 121,6 nanometri nell'ultravioletto lontano. Ciò corrisponde a uno spostamento verso il rosso di 2,66 (ricavato sottraendo 121,6 da 445,1 e dividendo per 121,6). Lo spettro mostra che le emissioni dall'ossigeno VI, dall'azoto V e dal carbonio IV sono spostate verso il rosso della stessa entità. I numeri romani sono di un'unità superiori al numero di elettroni strappati agli atomi.**

**Il valore 3,53 è il massimo spostamento verso il rosso rilevato di qualsiasi oggetto celeste conosciuto e appartiene al quasar OQ 172. Il suo spettro, ottenuto con il telescopio da quattro metri di Cerro Tololo, mostra la riga Lyman-alfa spostata da una lunghezza d'onda di 121,6 nanometri a una lunghezza d'onda di 550,8 nanometri nella regione verde dello spettro visibile. Le numerose righe di assorbimento sulla sinistra della riga Lyman - alfa dell'idrogeno ionizzato sono causate da nubi di gas che si trovano lungo la linea visuale tra il quasar e il sistema solare.**

Successivamente Smith esplorò una lunga striscia del cielo meridionale con il telescopio Curtis Schmidt. Inoltre, insieme ad Arthur A. Hoag del Kitt Peak National Observatory estese l'esplorazione a grandezze inferiori in campi ristretti del rilevamento Schmidt inserendo un reticolo di diffrazione a trasmissione (equivalente a un prisma) nel fuoco primario del telescopio da quattro metri di Cerro Tololo. Io continuai a esaminare i quasar non rivelati nei due rilevamenti effettuati con il più sensibile spettrometro Vidicon. L'insieme delle osservazioni costituisce la base dei nuovi risultati che descriverò.

Gli spettri dei quasar sono completamente diversi da quelli di tutti gli altri oggetti astronomici. Il grande spostamento verso il rosso rende visibili regioni dello spettro dell'ultravioletto lontano mai registrate in precedenza con i telescopi con base a terra. Il valore dello spostamento verso il rosso, spesso indicato con $Z$, si ottiene sottraendo la lunghezza d'onda di riposo (la lunghezza d'onda non spostata verso il rosso) di una riga di emissione dalla sua lunghezza d'onda osservata e dividendo la differenza per la lunghezza d'onda di riposo. L'aspetto che più colpisce negli spettri dei quasar è la riga Lyman-alfa dell'idrogeno atomico, spostata anche di un fattore 4,5 dalla sua posizione normale a una lunghezza d'onda di 121,6 nanometri nell'ultravioletto fino a circa 550 nanometri, la lunghezza d'onda della luce giallo-verdastra. In tal caso $Z$ risulterebbe 3,5 (550 meno 121,6 diviso per 121,6). Negli spettri dei quasar con grandi spostamenti verso il rosso sono pure intense le righe di emissione dell'ossigeno, dell'azoto e del carbonio con normali lunghezze d'onda ultraviolette comprese tra 103,4 e 154,9 nanometri.

Inoltre, in molti casi le righe di emissione sono larghe, indicando che una parte del gas che circonda il quasar è in moto a velocità fino a 10 000 chilometri al secondo. Le condizioni fisiche dedotte dalle intensità delle varie righe mostrano che il gas è più caldo del gas nelle normali nebulose e che la sorgente centrale del quasar non irraggia affatto come una stella normale. Il punto principale della questione consiste però nel fatto che la riga Lyman-alfa è la caratteristica più intensa degli spettri dei quasar ed è perciò quella più facilmente rivelata sulle lastre di Schmidt eseguite attraverso un prisma obiettivo. Pertanto il metodo del prisma obiettivo favorisce il rilevamento di quasar con grandi spostamenti verso il rosso.

Se gli spostamenti verso il rosso vengono interpretati come velocità di recessione, lo spostamento verso il rosso necessario per muovere la radiazione Lyman-alfa di 364,8 nanometri nella regione dello spettro dell'ultravioletto vicino, equivalente a uno spostamento verso il rosso pari a 2, corrisponde all'80 per cento della velocità della luce. OQ 172, il quasar con il più grande spostamento verso il rosso che si conosca, 3,53, si sta chiaramente allontanando al 91 per cento della velocità della luce. Nella stessa scala le stelle

all'interno delle galassie hanno velocità pari allo 0,1 per cento della velocità della luce e le galassie vicine si stanno allontanando da noi a non più dell'1 per cento della velocità della luce.

In più di un quarto di secolo di misurazioni effettuate sugli spostamenti verso il rosso delle galassie con il telescopio da 100 pollici di Mount Wilson e con il telescopio da 200 pollici di Monte Palomar, il massimo valore registrato da Milton L. Humason è stato uno spostamento verso il rosso di 0,2, ottenuto nel 1949. Passarono altri 11 anni prima che Rudolph Minkowski, sempre con il telescopio da 200 pollici, spingesse il limite fino a 0,46, un record che resistette per molti anni. Entro due anni dalla scoperta di Maarten Schmidt del 1963 del primo quasar, 3C 273, dotato di uno spostamento verso il rosso piuttosto modesto, 0,158, la «barriera» 2 nello spostamento verso il rosso venne infranta. Il valore massimo di 3,53 è stato registrato nel 1973.

Se si interpretano gli spostamenti verso il rosso dei quasar come dovuti all'espansione dell'universo, essi non sono oggetti vicini in moto ad alta velocità, ma sono oggetti estremamente lontani. Questa è una conseguenza diretta della conclusione cui giunse Edwin P. Hubble nel 1929, soprattutto sulla base delle osservazioni di Humason e secondo la quale le galassie si stanno allontanando da noi (e una dall'altra) a velocità proporzionali alle loro distanze. Sulla base della scala attuale dell'universo una galassia con uno spostamento verso il rosso pari a solo 0,01 (ossia a una velocità apparente di recessione di 3000 chilometri al secondo) dovrebbe trovarsi a una distanza di 200 milioni di anni luce. Si tratta già di una distanza notevole; la galassia a spirale più vicina, la galassia di Andromeda, dista soltanto circa due milioni di anni luce. I quasar sono di gran lunga più lontani e quelli con i maggiori spostamenti verso il rosso si avvicinano all'orizzonte, ossia al limite, dell'universo osservabile.

La misurazione delle distanze ha sempre costituito un problema per l'astronomia. Solo le stelle più vicine sono state veramente misurate con metodi trigonometrici, come un geometra misura le distanze sulla Terra; tutte le misurazioni al di là di queste stelle sono state estrapola-

**Questa rappresentazione schematica dell'universo in differenti epoche illustra come i quasar possano fornire spiegazioni sulle condizioni esistenti in epoche antiche. La luce che arriva oggi alla Terra da un quasar con uno spostamento verso il rosso pari a 3 è stata emessa circa 15 miliardi di anni fa, soltanto cinque miliardi di anni dopo l'ipotetico big bang che ha avviato l'espansione dell'universo da uno stato denso ($t = 0$). All'età di cinque miliardi di anni l'universo aveva dimensioni pari solo a un quarto di quelle attuali. I cerchi in colore indicano la distanza coperta dalla radiazione del quasar mentre l'universo si espande e il quasar e la nostra galassia si allontanano. La radiazione raggiunge alla fine i telescopi sulla Terra.**

te. Nessuno sostiene che gli spostamenti verso il rosso rivelino le distanze assolute delle galassie, ma si è generalmente d'accordo sul fatto che gli spostamenti verso il rosso costituiscono una buona misura delle distanze relative. Il concetto di universo in espansione implica che un oggetto con un grande spostamento verso il rosso viene visto come era molto tempo fa. La luce che arriva a noi da un quasar con uno spostamento verso il rosso pari a 3 è stata emessa circa 15 miliardi di anni fa. La luce proveniente da un quasar con uno spostamento verso il rosso pari solo a 1 ha viaggiato per 10 miliardi di anni, ossia la metà dell'età dell'universo.

Un'ulteriore conseguenza di un universo in espansione è che lo spostamento verso il rosso fornisce una misura diretta del fattore di espansione. Quando è stata emessa la luce da un quasar con uno spostamento verso il rosso pari a 3, l'universo aveva dimensioni pari soltanto a un quarto di quelle attuali, cioè tutto era quattro volte più ravvicinato. La distribuzione degli spostamenti verso il rosso dei quasar fornisce perciò molte indicazioni sulla struttura e sulle caratteristiche del primo universo. Aggiungo subito che i particolari delle relazioni matematiche tra gli spostamenti verso il rosso osservati e le proprietà dell'universo sono alquanto complessi e certamente controversi, date le labili ipotesi che si possono ricavare da dati ancora limitati ottenuti dalle osservazioni. Ciononostante, certi aspetti dell'indagine sono relativamente indipendenti dai minuti particolari e noi concentreremo su di essi l'attenzione.

Nei primi anni di ricerca sui quasar molti astronomi speravano che questi oggetti aiutassero a decidere quale delle molte ipotesi possibili sull'evoluzione dell'universo fosse corretta. Tale speranza non fu soddisfatta, ma verso la fine degli anni sessanta Maarten Schmidt scoprì una notevole proprietà dei quasar lontani: essi sono molto più numerosi nello spazio a grandi distanze di quanto lo siano nelle nostre vicinanze. In corrispondenza a uno spostamento verso il rosso di 2, pari a circa 13 miliardi di anni fa, la loro densità era 1000 volte superiore di quella attuale. Evidentemente, qualunque sia stato il processo che ha dato origine ai quasar, esso deve essere stato estremamente attivo quando l'universo era giovane mentre oggi si è praticamente ridotto a nulla.

All'epoca del lavoro compiuto da Maarten Schmidt sulla densità spaziale dei quasar il massimo spostamento verso il rosso conosciuto era pari a 2,88. Considerata l'enormità di tale valore in confronto alle previsioni di solo qualche anno addietro, questo risultato suggerì che i quasar con spostamenti verso il rosso molto maggiori fossero abbondanti. Perché non venivano scoperti quasar con spostamenti verso il rosso pari a 3? Sia Schmidt sia Sandage richiamarono l'attenzione su questo punto e avanzarono l'ipotesi che gli spostamenti verso il rosso dei quasar avessero un limite.

Le implicazioni di un tale limite erano estremamente importanti. Il limite comportava che almeno per un tipo di oggetti gli astronomi dovessero cercare ai confini dell'universo. Pareva quasi che i quasar si fossero improvvisamente formati in una grande esplosione di attività circa 15 miliardi di anni fa. Una tale idea, ovviamente, era destinata a influenzare notevolmente le ipotesi sulla natura fondamentale dei quasar. Se si suppone anche che i quasar rappresentino enormi processi energetici nel centro delle galassie, il limite allo spostamento verso il rosso ha importanti implicazioni per l'evoluzione delle galassie stesse.

Naturalmente gli astronomi continuarono a cercare quasar con spostamenti verso il rosso sempre maggiori. Un'ovvia deduzione fu che il limite apparente potesse essere il risultato di qualche effetto di selezione nei metodi di identificazione, anche se i quasar conosciuti erano stati scoperti con esplorazioni del tutto eterogenee basate su metodi diversi. Era controverso se i quasar con spostamenti verso il rosso maggiore di 2,5 o di 3 potessero continuare a essere più luminosi nell'ultravioletto delle stelle ordinarie. Il quasar con uno spostamento verso il rosso di 2,88 non presentava una particolare intensità nell'ultravioletto e, dal momento che la sua riga più intensa di emissione, la Lyman-alfa, è spostata verso il centro dello spettro visibile, risulterebbe difficile distinguere questo quasar e altri oggetti simili dalle stelle normali.

In realtà, quando nel 1973 venne scoperto il quasar con uno spostamento verso il rosso di 3,53, esso si rivelò di colore rosso. Venne identificato solo perché i radioastronomi avevano determinato la sua posizione con elevata precisione. Se si fosse trattato di un quasar privo di segnali radio, non avrebbe destato attenzione. La sua scoperta, insieme ai dubbi sulla completezza delle precedenti ricerche, mantenne aperta la questione sulla possibile esistenza di un limite allo spostamento verso il rosso.

**La densità dei quasar 15 miliardi di anni fa era più di cento volte superiore a quella attuale, secondo i calcoli eseguiti da Maarten Schmidt, che scoprì il primo quasar nel 1963. L'assenza di quasar con spostamento verso il rosso maggiore di 3,53 indica che gli astronomi hanno esplorato la più remota epoca di formazione dei quasar. A tutt'oggi si conoscono all'incirca 1500 quasar.**

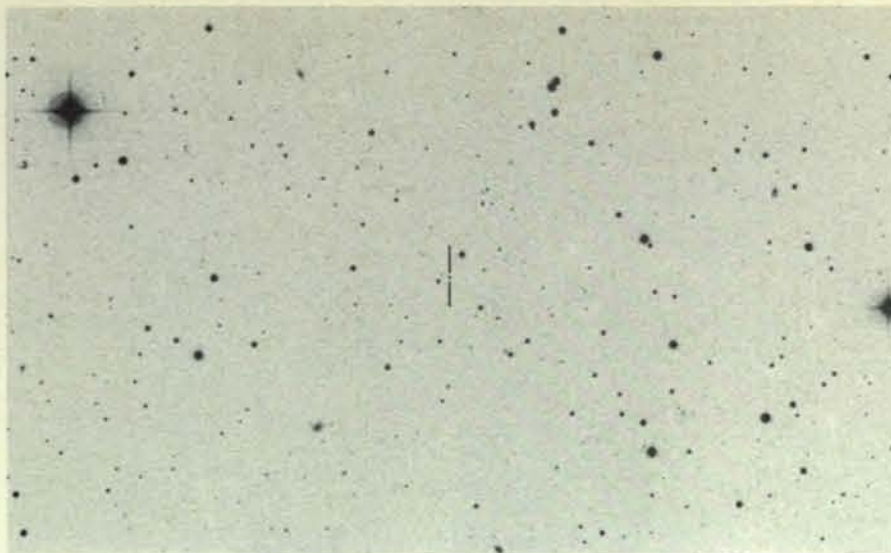
**La fotografia, eseguita con il telescopio da quattro metri di Cerro Tololo, mostra l'ammasso di galassie del Centauro, uno dei molti ammassi noti. Distante circa 225 milioni di anni luce, esso comprende circa 250 grandi galassie, separate tra loro di circa 700 000 anni luce. La comprensione del processo di formazione degli ammassi dovrebbe aiutare a spiegare come si sono formate le galassie nel primo universo.**

Il metodo del prisma obiettivo da noi messo a punto a Cerro Tololo offrì nuove possibilità di affrontare il problema. Dato che il metodo si basava soltanto sul rilevamento delle righe di emissione, esso era completamente indipendente dal colore del quasar, ultravioletto o altro. Anche se il metodo aveva inevitabilmente dei limiti, questi erano almeno differenti da quelli di tutte le altre ricerche.

Il metodo del prisma obiettivo consentì di rilevare la riga Lyman-alfa dei quasar fino all'estremità rossa dello spettro come permettono le emulsioni fotografiche. L'emulsione scelta da Hoag e Smith per la loro approfondita esplorazione con il telescopio da quattro metri era sensibile alla riga Lyman-alfa fino a 690 nanometri nell'estremo rosso, equivalente a uno spostamento verso il rosso di 4,7. Ciononostante, nessuno dei 71 nuovi quasar scoperti nel corso della loro esplorazione rivelò uno spostamento verso il rosso maggiore di 3,45, che era vicino, ma non superiore al precedente limite. Colpiti dall'importanza del risultato, Robert F. Carswell dell'Università di Cambridge e Smith dimostrarono successivamente che l'esplorazione di Hoag e Smith non aveva una sensibilità sufficiente, per spostamenti verso il rosso superiori a 3,5, perché la loro densità effettivamente diminuisse. Si imponeva perciò una nuova esplorazione.

Eseguii osservazioni con il telescopio da quattro metri di Cerro Tololo in un programma ottimizzato per la ricerca di quasar con spostamenti verso il rosso compresi tra 3,7 e 4,7, nel caso fossero esistiti. Lavorando con un reticolo a trasmissione di sensibilità più elevata nella regione di 600 nanometri insieme a un filtro per eliminare la luce del cielo notturno a lunghezze d'onda inferiori a 570 nanometri, fui in grado di rivelare quasar tre volte più deboli di quanto avevano fatto Hoag e Smith nella regione critica compresa tra 570 e 690 nanometri. Scoprii 15 oggetti con righe di emissione, ciascuno dei quali avrebbe potuto trovarsi al di sopra del limite di 3,5. Successive osservazioni con lo spettrometro Vidicon mostrarono, però, che in nessun caso la riga rilevata sulla lastra fotografica era quella Lyman-alfa. In cinque oggetti la riga risultò dovuta al carbonio contenuto nei quasar con spostamenti verso il rosso compresi tra 2,8 e 3,4; tutti gli oggetti rimanenti erano o quasar con spostamenti verso il rosso prossimi a 1 o galassie con spostamenti verso il rosso di circa 0,2. Pur essendo in un certo senso deludenti questi risultati facevano sperare di poter osservare, se esistevano, quasar con spostamenti verso il rosso pari a 4. Tutti i nuovi quasar furono scoperti dal rilevamento di righe più deboli della riga Lyman-alfa, completamente all'opposto delle precedenti esplorazioni eseguite con il metodo del prisma obiettivo.

La mia successiva analisi quantitativa dimostrò che i nuovi risultati del rilevamento potevano essere realmente compresi solo se la densità spaziale di quasar al di là di uno spostamento verso il rosso di 3,5 era di un fattore tre o più al di sotto del valore osservato a uno spostamento verso il rosso di 3. Il limite non esclude che si possano trovare alcuni quasar a distanze maggiori, ma dimostra in modo convincente che abbiamo raggiunto il punto di inversione.

Valide prove a sostegno dell'esistenza di un limite di spostamento verso il rosso sono fornite dall'insuccesso dei radioastronomi nello scoprire quasar più lontani, anche se essi dispongono oggi di posi-

**Il più lontano oggetto, finora individuato, sulla base del suo spostamento verso il rosso di 3,53, è il quasar OQ 172, che si presenta come una comune stella in questa fotografia eseguita con il telescopio Schmidt da 1,2 metri di Monte Palomar. OQ 172 fu scoperto nel 1973 al Lick Observatory da E. Joseph Wampler, Lloyd B. Robinson, Jack Baldwin ed E. Margaret Burbidge.**

zioni molto precise che consentono di ignorare il colore dei possibili quasar. Recentemente gli astronomi che lavorano nel campo dei raggi X hanno fornito i dati raccolti dal satellite per raggi X, l'Osservatorio Einstein. Sebbene il telescopio a raggi X del satellite abbia facilmente ripreso OQ 172, il quasar più lontano, e molti altri quasar con spostamenti verso il rosso di circa 3,1, non si è scoperto alcun oggetto che superi tale record. È probabile che le ricerche progrediscano con lo Space Telescope, la cui messa in orbita è prevista per il 1985.

Le nostre ricerche a Cerro Tololo indicano che la densità dei quasar nello spazio continua a mantenersi elevata fino a spostamenti verso il rosso di 3,2. Il limite di 3,5 implicherebbe una brusca variazione delle proprietà dell'universo. Una delle spiegazioni più semplici è che i quasar si siano formati improvvisamente circa 15 miliardi di anni fa, un evento senza dubbio considerevole nell'evoluzione dell'universo. In alternativa, potrebbe esserci uno schermo assorbente di polvere o di qualche altro tipo di materiale in corrispondenza di uno spostamento verso il rosso di 3,5 capace di impedire l'osservazione dei quasar più lontani. Sarebbe estremamente notevole anche il fatto che l'universo fosse trasparente da una parte del limite e opaco dall'altra. In entrambi i casi è probabile che ulteriori ricerche sull'argomento conducano a importanti risultati nei prossimi anni.

Finora mi sono occupato soltanto della distribuzione radiale dei quasar, che è il punto di vista storico lungo il quale si è sviluppato il loro studio. Ora che gli esemplari di quasar di Cerro Tololo sono stati completamente analizzati, la base dei dati relativi a grandi spostamenti verso il rosso è sufficiente per potersi occupare della distribuzione tridimensionale dei quasar. Combinando gli spostamenti verso il rosso con le posizioni dei quasar nel cielo si può costruire un quadro tridimensionale dei quasar nello spazio. Queste sono le sole informazioni disponibili su come era la struttura dell'universo da 13 a 15 miliardi di anni fa. La distribuzione tridimensionale può anche fornire importanti informazioni sulla natura dei quasar stessi.

Se si osserva la distribuzione delle galassie in fotografie che riprendono grandi regioni del cielo, appare evidente che le galassie sono tutt'altro che uniformemente distribuite. Esse appaiono in coppie, in piccoli gruppi, in gruppi più grandi e in grandi ammassi. Spesso vi sono regioni vuote con poche galassie o addirittura nessuna. Lo studio della distribuzione galattica è fondamentale per le teorie sulla formazione delle galassie. I dati attualmente disponibili indicano che nei primissimi stadi dell'universo la materia era uniformemente distribuita in forma gassosa. Come poterono, dunque, originarsi le condensazioni dalle quali si formarono in seguito nel corso dell'espansione dell'universo le stelle, le galassie e gli ammassi di galassie?

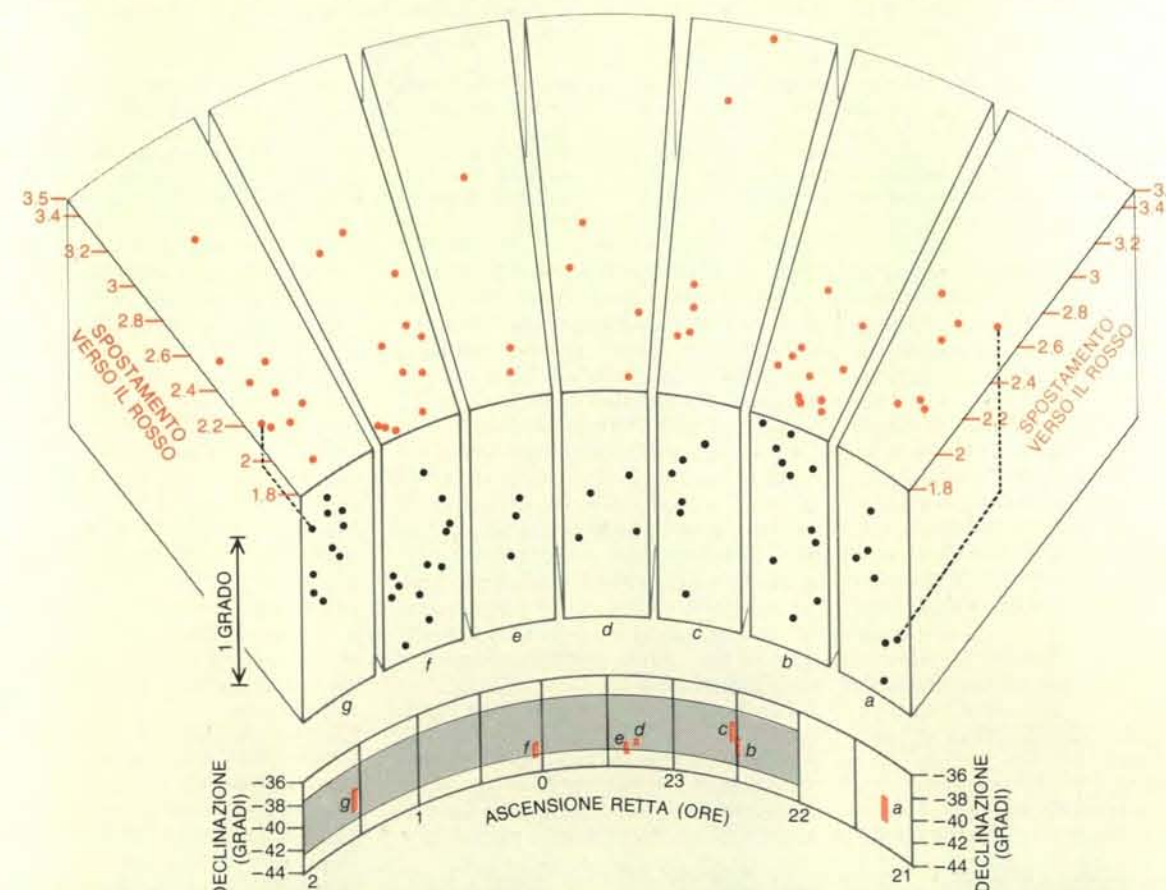
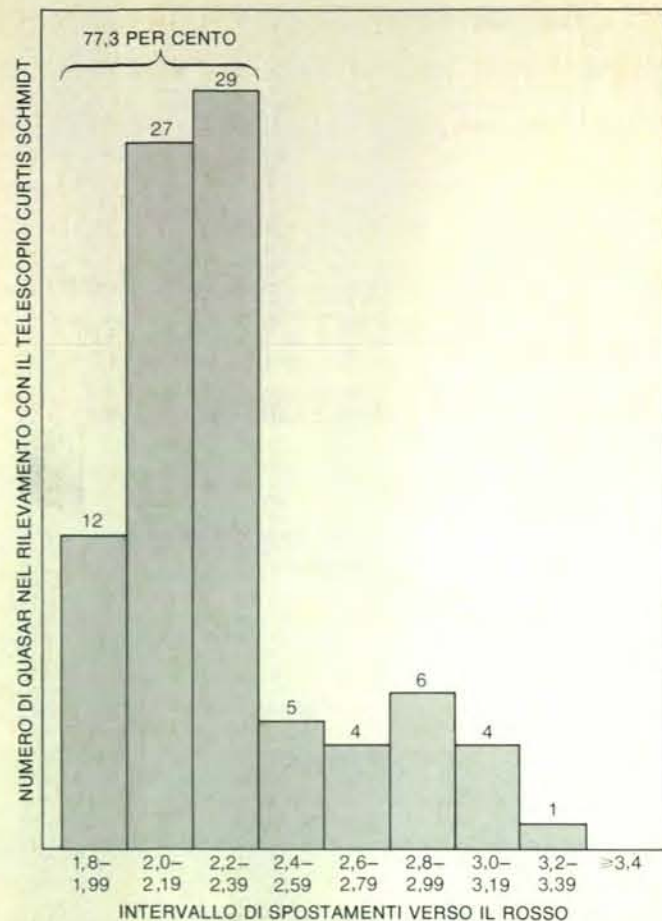
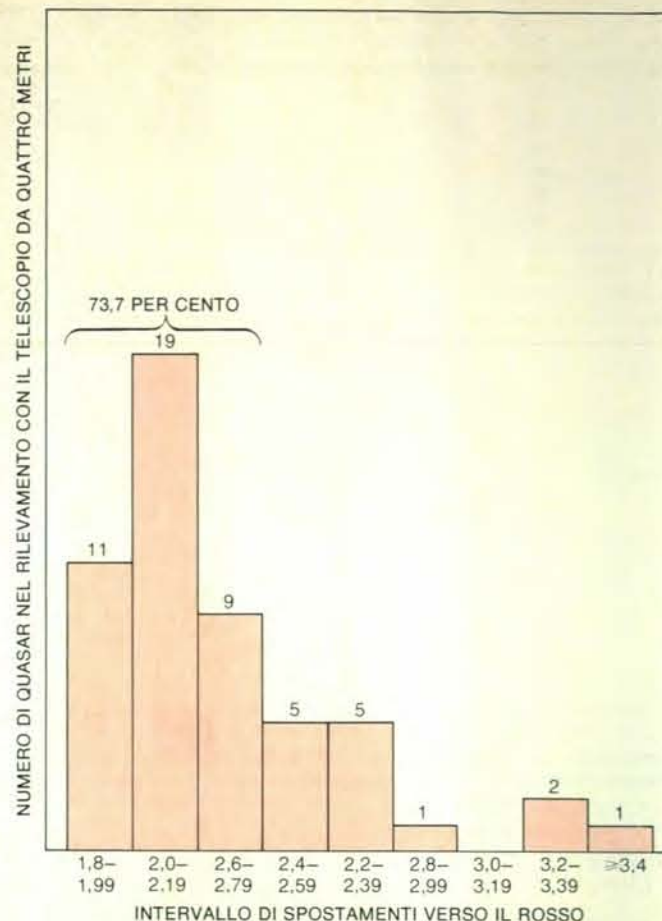
Una volta apparsa una condensazione con autogravità sufficiente a opporsi all'espansione dell'universo, si possono immaginare vari momenti nei quali essa potrebbe collassare sotto l'azione del suo stesso peso, formando una stella o addirittura un'intera galassia di stelle. La maggior parte dei calcoli dimostra che una volta che un corpo gassoso ha iniziato il collasso la caduta di materia verso il centro è immediata e porta a un rapido aumento della densità. Se il gas che cade verso il centro è di dimensioni galattiche, è facile immaginare che un collasso di materia in fuga potrebbe creare un quasar nel centro della galassia come processo dell'evoluzione della galassia. È questa una possibile correlazione tra i quasar e la formazione delle galassie. Un'altra possibilità è che i quasar siano per loro natura sufficientemente potenti da influenzare processi di formazione su una scala decisamente maggiore di quella di una singola galassia. In tal caso la distribuzione dei quasar potrebbe essere diversa dalla distribuzione delle galassie, che possiamo osservare soltanto in un tempo molto ritardato dell'espansione dell'universo.

Per ricavare una prima impressione sulla distribuzione dei quasar analizziamo i dati di Cerro Tololo sotto diversi aspetti rappresentativi. La regione del cielo analizzata nelle esplorazioni di Cerro Tololo copre una banda compresa tra 37,5 e 42,5 gradi di latitudine celeste sud che si estende per un quarto della sfera celeste *(si veda l'illustrazione in basso in questa pagina)*. La banda è stata ripresa in una sequenza di circa 15 lastre leggermente sovrapposte eseguite con il telescopio Curtis Schmidt, ciascuna delle quali copriva un campo di cinque gradi di lato. (Le due stelle all'estremità del Gran Carro che individuano la direzione della Stella Polare hanno all'incirca la stessa distanza angolare.) Il grande riflettore da quattro metri eseguì esposizioni in profondità in sette piccole regioni, ciascuna di circa un grado quadrato, all'interno dell'ampia banda fotografata con il Curtis Schmidt. In realtà il telescopio da quattro metri campionò lunghi e sottili tunnel dello spazio.

Il rilevamento effettuato con il Curtis Schmidt rivelò 88 quasar con spostamenti verso il rosso maggiori o uguali a 1,88. In

**Le ricerche di quasar a Cerro Tololo sono state eseguite lontano dal piano della Via Lattea per evitare l'oscuramento della polvere interstellare o di qualche altro materiale che avrebbe assorbito la luce proveniente da galassie e quasar lontani. La banda grigia indica la regione di 340 gradi quadrati esplorata dalla ricerca di Smith con il telescopio Curtis Schmidt da 60 centimetri munito di un prisma disposto davanti all'apertura. Le sette piccole zone in colore, che coprono complessivamente 5,1 gradi quadrati, sono state esplorate da Smith e da Arthur A. Hoag con il telescopio da quattro metri. Il telescopio da quattro metri era in grado di registrare oggetti circa quattro volte più deboli di qualsiasi oggetto rivelato dal telescopio Curtis Schmidt.**



**Il rilevamento eseguito con il telescopio Curtis Schmidt individuò 88 quasar con uno spostamento verso il rosso maggiore di 1,8. Dato che gli spostamenti verso il rosso sono correlati alla distanza, gli 88 oggetti si possono rappresentare graficamente come se fossero distribuiti in un volume di spazio tridimensionale. I puntini in nero nella parte anteriore del diagramma indicano la posizione dei quasar nel cielo, mentre i puntini in colore sulla superficie superiore mostrano lo spostamento verso il rosso corrispondente a ciascun quasar. Lo schema di rappresentazione grafica è indicato dalle linee tratteggiate che collegano il quasar all'estrema destra con il suo spostamento verso il rosso di 3,16 e il quasar all'estrema sinistra con il suo spostamento verso il rosso di 1,96. Per chiarezza sono state omesse altre linee di collegamento. La dimensione verticale della regione esplorata è stata ingrandita di un fattore due rispetto a quella orizzontale per evidenziare la distanza tra i quasar.**

**Il rilevamento eseguito con il telescopio da quattro metri ha individuato 53 quasar con uno spostamento verso il rosso maggiore di 1,8, tra i quali sei quasar già scoperti nell'esplorazione con il telescopio Curtis Schmidt. Qui le sette piccole zone esplorate con il grande telescopio sono state arbitrariamente dilatate in forma di cunei di volume uniforme. I puntini in nero sulla faccia anteriore dei cunei mantengono comunque la esatta posizione di ogni quasar all'interno della propria area di esplorazione. Sono evidenti alcune coppie e piccoli gruppi di quasar con spostamenti verso il rosso simili. Per esempio, nella zona *f*, sono raggruppati vicino a uno spostamento verso il rosso di 1,84 quattro quasar a una distanza di circa 12 miliardi di anni luce. Sono tutti contenuti all'interno di un volume esteso per non più di 200 milioni di anni luce.**

**La distribuzione dei quasar secondo il loro spostamento verso il rosso indica che circa i tre quarti dei quasar nelle due esplorazioni di Cerro Tololo hanno spostamenti verso il rosso compresi tra 1,8 e 2,4. La distribuzione degli 88 quasar dell'esplorazione Curtis Schmidt è rappresentata a sinistra, mentre quella dei 53 quasar rilevati nell'esplorazione con il telescopio da quattro metri è rappresentata a destra. L'esplorazione Curtis Schmidt ha individuato un quasar per ogni quattro gradi quadrati di cielo. L'esplorazione con il telescopio da quattro metri, capace di rilevare oggetti più deboli, ha individuato poco più di 10 quasar per grado quadrato esplorato. I quasar rilevati nell'esplorazione con il telescopio da quattro metri, pur essendo in genere meno luminosi di quelli dell'esplorazione Curtis Schmidt, non erano più lontani.**

una regione del cielo molto meno estesa il rilevamento con il telescopio da quattro metri identificò 53 quasar con spostamenti analoghi verso il rosso, tra i quali sei scoperti anche nel rilevamento effettuato con il telescopio Schmidt. (Un piccolo numero dei quasar più luminosi identificati in entrambe le ricerche è stato scoperto prima da altri astronomi.) Il rilevamento Curtis Schmidt evidenziò un quasar con grande spostamento verso il rosso all'incirca ogni quattro gradi quadrati di cielo; il rilevamento con il telescopio da quattro metri rivelò all'incirca un numero di quasar 40 volte superiore per unità di superficie esplorata, pari a circa 10 quasar con grande spostamento verso il rosso ogni grado quadrato.

La spiegazione della maggiore frequenza di scoperte effettuate con il sistema da quattro metri va ricercata ovviamente nel fatto che esso poteva rivelare oggetti molto più deboli. Il rilevamento con il telescopio da quattro metri fino a una grandezza limite di 21, rivelando quindi oggetti di circa 1,5 grandezze, ossia quattro volte, più deboli di quelli che potevano essere registrati con il piccolo telescopio Schmidt. Anche se i quasar scoperti con il rilevamento con il telescopio da quattro metri sono generalmente più deboli non mostrano generalmente maggiori spostamenti verso il rosso. La maggior parte dei 135 quasar dei due rilevamenti mostrano spostamenti verso il rosso prossimi a 2. Ciò era previsto sia perché il metodo di ricerca favorisce la scoperta di tali quasar sia perché i quasar con tali spostamenti verso il rosso sono chiaramente i più abbondanti. I 135 quasar esaminati a Cerro Tololo costituiscono circa la quarta parte di tutti i quasar scoperti finora con spostamento verso il rosso superiore a 1,8.

Per dare un'idea della effettiva distribuzione dei quasar nello spazio, ho rappresentato graficamente gli 88 quasar del rilevamento Schmidt su un solido tridimensionale a forma di torta *(si veda l'illustrazione in alto nella pagina precedente)*. La faccia frontale del solido è un arco che localizza gli 88 quasar secondo la declinazione, o latitudine, e l'ascensione retta (l'equivalente astronomico della longitudine, che suddivide la sfera celeste in 24 spicchi, ciascuno largo 15 gradi). Lo spostamento verso il rosso di ciascun quasar è rappresentato radialmente nella parte superiore del solido su una scala che va da uno spostamento verso il rosso di 1,8, a uno spostamento verso il rosso di 3,5.

Questa rappresentazione grafica rivela che in una regione compresa tra le due e le quattro ore di ascensione retta i quasar appaiono più rari di quanto lo siano altrove. Se la scarsità riflette una reale disuniformità dei quasar in quella parte dello spazio, si tratterebbe di un fatto sorprendente. Per tale motivo i miei colleghi e io preferiamo pensare per ora che l'apparente scarsità sia dovuta a un effetto selettivo nel processo di esplorazione. Se guardiamo la distribuzione dei quasar su piccola scala, si possono vedere casi nei quali coppie e piccoli gruppi di quasar sono molto vicini tra loro. Il raggruppamento più evidente si manifesta quando si rappresentano graficamente in un analogo modo tridimensionale i quasar rilevati con il telescopio da quattro metri *(si veda l'illustrazione in basso nella pagina precedente)*. Un gruppo di quattro quasar con spostamenti verso il rosso compresi tra 1,83 e 1,86 è ammassato in una regione non più ampia di 200 milioni di anni luce nei pressi del punto con ascensione retta di zero ore.

Questi raggruppamenti sono interessanti perché sono all'incirca delle stesse dimensioni dei superammassi di galassie distribuiti qua e là nello spazio circostante. I superammassi sono regioni con diametro di circa 300 milioni di anni luce, le quali ospitano numerosi ammassi di galassie. Si tratta delle più grandi strutture conosciute dell'universo. Finora non si sono mai osservati superammassi con spostamenti verso il rosso che raggiungano il valore 2. Il gruppo di quasar con spostamenti verso il rosso attorno a 1,84 potrebbe mascherare l'esistenza di un lontanissimo superammasso? Forse sarà in grado di dircelo lo Space Telescope.

Considerata l'importanza dell'analisi visiva dei dati, la questione della distribuzione dei quasar finisce con l'imporre un'analisi statistica. Esiste uno schema fondamentale di distribuzione? I raggruppamenti osservati sono semplicemente fluttuazioni casuali? Dopo tutto, se spargiamo sul pavimento dei chicchi di riso, alcuni cadranno più vicini della distanza media tra i chicchi. Negli ultimi anni sono state messe a punto potenti tecniche statistiche da applicare ai rilevamenti galattici per indagare proprio su tali questioni. Con le opportune modifiche che tengano conto dei grandi spostamenti verso il rosso e degli effetti selettivi del metodo di rilevamento, queste tecniche si possono applicare anche ai quasar.

L'obiettivo fondamentale è quello di scoprire se i quasar presentano qualche scostamento da una distribuzione uniforme e casuale. Se è così, si vorrebbe sapere se i quasar si ammassano e quale forma assumono gli ammassi. In alternativa, si potrebbe scoprire che i quasar sono contrari alla formazione di ammassi e si trovano più lontani del previsto. Sarebbe così se la presenza di un quasar in un dato luogo dovesse impedire la formazione di altri quasar nelle vicinanze. La formulazione matematica di tali possibilità e le verifiche per evidenziarle non ci riguardano necessariamente in questa sede, ma le principali possibilità conducono esse stesse a una rappresentazione grafica bidimensionale. Sia una forte tendenza ad ammassarsi sia una forte tendenza opposta sono riconoscibili a prima vista. Quando, però, entrambe le tendenze sono deboli, è difficile distinguere la distribuzione da una distribuzione effettivamente uniforme e casuale *(si veda l'illustrazione nella pagina successiva)*.

Posso riassumere i risultati ottenuti finora dalle esplorazioni di Cerro Tololo dicendo che l'analisi statistica è coerente con l'esistenza di una distribuzione dei quasar uniforme e casuale. Le coppie e i gruppi visibili a occhio non sono evidentemente altro che fluttuazioni casuali. Per i più di 30 quasar con spostamenti verso il rosso compreso tra 1,8 e 2,2 nel campione osservato con il telescopio da quattro metri, la minima distanza media è di 400 milioni di anni luce, ben al di sotto del limite previsto per una distribuzione casuale. Per i quasar con spostamenti verso il rosso maggiori, per i quali le distanze medie si misurano in miliardi di anni luce,

si possono trarre analoghe conclusioni. Questi risultati parlano a favore dell'ipotesi, generalmente sostenuta ma difficile da confermare, che su grande scala l'universo sia omogeneo. Tale ipotesi è un requisito fondamentale per gli attuali modelli dell'universo.

Benché convincenti questi risultati costituiscono soltanto il primo passo di una ricerca continua che è probabile progredisca rapidamente quando verranno scoperti altri quasar. Se si estrapola l'ammassarsi delle galassie osservato all'epoca e alla scala coperta nel rilevamento dei quasar, si può dimostrare che l'ammassarsi delle galassie non si dovrebbe poter rilevare nei campioni esistenti. Sarà interessante vedere se lo Space Telescope potrà migliorare la sensibilità delle verifiche della distribuzione fino al punto di poter rivelare l'ammassarsi delle galassie. La previsione è che esistano ammassi di galassie a uno spostamento verso il rosso pari a 2, ma che essi risaltino sul fondo di materia più densa in tale epoca con minor evidenza di quanto avvenga per gli ammassi con minori spostamenti verso il rosso osservati in un'epoca successiva in un universo molto più espanso e quindi meno denso. Al tempo stesso non si possono ignorare le coppie e i gruppi di quasar che appaiono evidenti nell'esplorazione fatta con il telescopio da quattro metri. Dopo tutto, una teoria della formazione delle galassie postula che esse abbiano avuto origine da fluttuazioni casuali di densità avvenute nell'universo primitivo. I raggruppamenti di quasar potrebbero rappresentare regioni analoghe di densità intensificata.



**Il quadro della distribuzione dei quasar viene analizzato perché esso potrebbe portare a una più completa conoscenza della formazione delle galassie nel primo universo. I sei schemi mostrati in questa illustrazione sono analoghi bidimensionali di varie distribuzioni possibili. Se i quasar sono fortemente ammassati *(a)* o se hanno invece una forte tendenza contraria all'ammasso *(b)*, la cosa dovrebbe essere ben evidente. Se mostrano solo una debole tendenza ad ammassarsi *(c)* o una debole tendenza opposta *(d)*, potrebbero essere necessarie approfondite verifiche statistiche per dimostrare che le distribuzioni non sono uniformi e casuali. Due simulazioni al calcolatore di distribuzioni uniformi e casuali *(e,f)* mostrano che l'occhio non è un giudice attendibile di casualità.**

In seguito ai risultati ottenuti a Cerro Tololo si può ora tentare di descrivere a cosa rassomigliava l'universo nel periodo compreso tra 12 e 15 miliardi di anni fa. Supponiamo che la nostra galassia, la Via Lattea, si trovasse all'interno del gruppo di quattro quasar che, con gli strumenti attuali, mostrano uno spostamento verso il rosso di circa 1,85. Supponendo che la vita umana si sia evoluta fin dall'inizio dell'universo, che cosa vedremmo di notte? Prima di tutto, la stessa Via Lattea dovrebbe essere molto più luminosa di quanto lo sia oggi, perché sarebbe costituita da una popolazione più numerosa di stelle calde e giovani. I quattro quasar del gruppo dovrebbero essere chiaramente visibili a occhio nudo come stelle luminose; infatti, si potrebbero vedere senza telescopio anche quasar esterni al gruppo distanti fino a 300 milioni di anni luce. Ovviamente i quasar sarebbero stati scoperti e considerati fenomeni strani all'inizio della storia dell'astronomia. Se poi essi siano o meno il nucleo luminoso di galassie si sarebbe accertato soltanto con l'avvento dei telescopi.

È bene ricordare che, per quanto plausibile possa essere la nostra attuale conoscenza sui quasar, esiste qualche probabilità che essa possa essere del tutto errata e una buona probabilità che essa sia errata in qualche importante particolare. Alcuni astronomi dubitano che i quasar siano distanti come i loro spostamenti verso il rosso indicano. Altri mettono in dubbio la realtà dell'elevata densità di quasar con spostamenti verso il rosso elevati e propongono che o i quasar fossero o più luminosi fin dall'inizio o che la loro intensità sia aumentata a causa del passaggio della loro radiazione nelle vicinanze di galassie che si trovano in mezzo e che agiscono come lenti gravitazionali. Raramente, se mai è accaduto, un grande corpo di dati viene raccolto e subito spiegato di colpo da una suggestiva teoria. E così è probabile che accada per i quasar. L'argomento sarebbe ben più povero senza le attuali ipotesi, per quanto audaci esse possano essere.

# TEMI METAMAGICI

*di Douglas R. Hofstadter*

## *Un articolo autoreferenziale sull'articolo dello scorso aprile sull'autoriferimento*

In questo articolo riprenderò il tema dell'autoriferimento che ho già trattato nel numero di aprile dello scorso anno. Vorrei approfittare di questo paragrafo d'apertura per avvertire i lettori non interessati all'argomento dell'autoriferimento che probabilmente vorranno smettere di leggere prima di raggiungere la fine di questo paragrafo, ossia questa proposizione, oppure questo inciso, anzi questa frase-nome, in breve, questo.

Dopo il mio articolo dell'aprile scorso ho ricevuto una grande quantità di posta autoreferenziale. Tony Durban astutamente insinuava: «Dato il grande volume di risposte che probabilmente avrà ricevuto, non penso che lei mi stia leggendo di persona». La lettera di John C. Waugh supplicava: «Aiuto, sono sepolta sotto una valanga di risposte di lettori». La sua lettera fu salvata e ora riposa in un mucchio molto ridotto, da cui pescherò per presentare qui quelle che maggiormente mi sono piaciute.

Prima di abbandonare l'argomento posta, vorrei riferire che il francobollo della cartolina di Ivan Vince, proveniente dall'Inghilterra, portava questa criptica dicitura: «Da indirizzarsi correttamente». Era un ordine dato dall'ufficio postale alla cartolina stessa? Se è così, le cartoline inglesi devono essere più intelligenti di quelle americane; non ho ancora incontrato una cartolina che sappia leggere, figuriamoci poi correggere il proprio indirizzo. (Mi è pervenuta una cartolina indirizzatami presso la rivista «Omni»!) Sono stato lusingato da un paio di complimenti autoneganti. Richard Ruttan ha scritto «Non so proprio dirle quanto mi sia piaciuto il suo primo articolo», e John Collins «Questa mia non riesce a esprimere quanto mi abbia deliziato l'articolo di aprile».

Un passatempo diffuso tra i miei corrispondenti è stato la ricerca di domande che contengono la propria risposta. Solo pochi sono riusciti a «saltar fuori dal sistema». Sembra che non sia facile aver successo in questa limitata forma artistica. John Flagg notava cinicamente (parafraso un po'): «Poni una domanda che si auto-risponde e otterrai una risposta che si auto-interroga». Una delle migliori era di Henry Taves: «Ricordo con piacere un esame scritto di storia in cui tra l'altro si chiedeva: "Scrivere una domanda adatta per l'esame finale di questo corso e rispondere". Per risposta non feci altro che copiare due volte quella frase». Mi è piaciuto molto, ma riflettendoci sopra ho iniziato a sospettare che ci fosse qualcosa che non andava. Voi cosa ne pensate?

Due strane domande ad auto-risposta mi sono state inviate da Richard Showstack: «Che domanda non verbo?» e «Qual è una domanda che menzioni la parola "ombrello" senza alcuna ragione apparente?» Jim Shiley mi ha mandato un'intelligente domanda che ho leggermente modificato in «Questa è una domanda retorica o è una domanda retorica?» Ancora di Shiley è il seguente contributo: «Prendete un foglio bianco e scrivete: "Fin dove arriverà nella pagina questa frase?" Ora, se qualche vostro amico poliglotta rileva che la stessa serie di fonemi in Uralo-Altaico significa "quindici centimetri", mandatemi un abbonamento a "Scientific American". Altrimenti, se la formulazione per iscritto di una domanda conta sia come domanda sia come unità di misura, avrò diritto almeno a un premio di consolazione. Ma penso di avere in qualche modo *violato* le regole».

Le mie proposte di soluzione al problema della domanda che può servire come propria risposta sono domande che più che *rispondersi* da sé si *provocano* da sé. Ad esempio: «Perché mi stai chiedendo *questo* senza malinconia?» È ovvio che quando la domanda è posta senza malinconia può provocare una risposta identica, con cui l'interlocutore manifesta il proprio stupore. Già, ma che cosa provoca *questa* domanda?

Philip Cohen riprende da Damon Knight questo aneddoto a proposito di una domanda che si risponde da sé: «Terry Carr, un nostro vecchio amico, ci inviò un indovinello su una cartolina, poi ci inviò la risposta su un'altra cartolina. In seguito ci mandò un altro indovinello: "Come si può tenere in sospeso un pollo?" senza farci avere la risposta. Dopo circa due settimane ci rendemmo conto che quella *era* la risposta.»

Molti dei veri capolavori inviatimi appartengono alla categoria di quelle che io chiamo le autodocumentazioni; un semplice esempio è il seguente: «Questa frase contiene dieci parole, ventisette sillabe e sessantasei lettere.» Anche la frase proposta da John Atkins è molto pulita: «"È di sedici lettere" lo è.» Queste forme di autodescrizione possono essere molto più involute e introspettive. Philip Cohen mi ha fornito un esempio di Howard Bergerson, un maestro dei giochi di parole: «In questa frase la parola E compare due volte, la parola TREDICI compare cinque volte, la parola CINQUE compare due volte, la parola DUE compare sette volte, la parola IN compare due volte, la parola COMPARE compare tredici volte, la parola FRASE compare due volte, la parola SETTE compare due volte, la parola LA compare tredici volte, la parola QUESTA compare due volte, la parola VOLTE compare tredici volte e la parola PAROLA compare tredici volte.»

Non è male, ma la medaglia d'oro per la categoria spetta a Lee Sallows, che si è sottoposto a questo *tour de force*: «Only the fool would take trouble to verify that his sentence was composed of ten a's, three b's, four c's, four d's, forty-six e's, sixteen f's, four g's, thirteen h's, fifteen i's, two k's, nine l's, four m's, twenty-five n's, twenty-four o's, five p's, sixteen r's, forty-one s's, thirty-seven t's, ten u's, eight v's, eight w's, four x's, eleven y's twenty-seven commas, twenty-three apostrophes, seven hyphens and, last but not least, a single!» (Solo il pazzo si prenderebbe la briga di verificare che la sua frase è composta da dieci a, tre b, quattro c, quattro d, quarantasei e, sedici f, quattro g, tredici h, quindici i, due k, nove l, quattro m, venticinque n, ventiquattro o, cinque p, sedici r, quarantun s, trentasette t, dieci u, otto v, otto w, quattro x, undici y, ventisette virgole, ventitre apostrofi, sette trattini e, ultimo ma non per importanza, un unico!) [Il traduttore italiano non è pazzo e quindi si limita a una traduzione letterale.]

Io (forse il pazzo) mi presi la briga di controllare tutto. Innanzitutto, però, feci qualche verifica di prova e devo dire che quando la prima verifica casuale si dimostrò positiva (mi pare che controllai il numero di *g*), la cosa ebbe un forte effetto psicologico: improvvisamente la frase mi divenne molto più credibile. Mi colpisce il fatto strano (e meraviglioso) che in certe situazioni la verifica di una piccola parte di una teoria può servire da potente rinforzo alla fiducia in una teoria. Forse è proprio qui il nocciolo della frase!

Sempre dello stesso tipo è il divertente rompicapo proposto dal noto logico Raphael Robinson. Il lettore deve completare questa frase: «In questa frase il numero di occorrenze di 0 è ____, di 1 è ____, di 2 è ____, di 3 è ____, di 4 è ____, di 5 è ____, di 6 è ____, di 7 è ____, di 8 è ____ e di 9 è ____». Gli spazi vuoti vanno riempiti con un numero di una o più cifre, scritto nella notazione decimale. Ci sono esattamente due soluzioni. I lettori potrebbero anche cercare due frasi che si descrivono l'un l'altra, o sequenze ancora più lunghe di questo tipo.

È chiaro che l'ultima parola in fatto di autodescrizione non spetterebbe a una frase che si limiti a fare l'inventario delle sue parti, bensì a una frase che includa anche una regola che dica come rimettere insieme quelle parti per creare l'intera frase - in breve, una frase che si autoriproduca. Una frase di questo genere è la ricostruzione in linguaggio ordinario fatta da Willard Van Orman Quine del classico omaggio metamatematico di Kurt Gödel a Epimenide il Cretese:

«comporta il falso quando segue la propria citazione» comporta il falso quando segue la propria citazione.

La frase di Quine dice in effetti al lettore come costruire una replica dell'enunciato e poi (per buona misura) aggiunge che la replica (non *essa stessa* per l'amor di Dio!) asserisce il falso! Viene in mente la famosa osservazione di Epilopside il Concretese (un cugino secondo di Epimenide) a Flora, una bella ragazza di cui non poteva ricambiare l'ardente amore (era fidanzato con la sorella gemella Fauna): «Fatti coraggio, mia cara, ho un suggerimento che ti può rallegrare. Basta prendere una di queste cellule del mio braccio e clonarla: ti troverai uno che mi assomiglia come un goccia d'acqua e pensa proprio come me. Ma stai attenta, pare che racconti delle grosse fandonie alle belle ragazze!».

Negli anni quaranta, John von Neumann lavorò a lungo al tentativo di progettare una macchina che potesse costruire una riproduzione di se stessa a partire dalle materie prime. Arrivò infine a un progetto teorico consistente di centinaia di migliaia di parti. Vista a posteriori e con un notevole grado di astrazione, l'idea che sta dietro la macchina autoriproducentesi di von Neumann rivela una forte somiglianza con il modo in cui si riproduce il DNA. E questo è a sua volta simile al metodo di Gödel per costruire una frase autoreferenziale in un linguaggio matematico nel quale, a prima vista, non sembra esserci modo di riferirsi al linguaggio stesso.

Ecco allora, per i lettori ambiziosi, il Gioco di von Neumann: creare una frase comprensibile e non irragionevolmente lunga che non solo elenchi le sue parti (a livello di parola o a livello di lettera), ma dica anche come metterle insieme in modo da ricostruire la frase stessa. (Si noti che il requisito che la frase sia «non irragionevolmente lunga» è diverso, molto diverso, da «ragionevolmente lunga».) L'elenco delle parti (o «seme») deve essere un inventario di parole o simboli tipografici, più o meno come nelle frasi create da Howard Bergerson e Lee Sallows. I simboli elencati devono essere chiaramente distinguibili dal testo che a essi si riferisce. Potrebbero per esempio essere racchiusi fra virgolette o stampati in un altro carattere o designati col loro nome. Non è importante la convenzione adottata, basta che sia chiara la distinzione. Il resto della frase (la «regola costruttiva») dovrebbe essere stampato normalmente, in quanto va visto non come materia prima tipografica, ma come insieme di istruzioni. Si tratta della distinzione tra uso e menzione di cui ho parlato lo scorso aprile, distinzione che non va trascurata se si vuole evitare una debolezza concettuale. (È questa una pecca della frase di Sallow che macchia la sua medaglia d'oro.)

La regola costruttiva non può riferirsi al materiale stampato normalmente, ma solo alle parti dell'inventario. La regola costruttiva non deve quindi in alcun modo riferirsi a se stessa! Inoltre la regola costruttiva deve definire la struttura in modo esplicito (ed è questo l'aspetto più sottile e probabilmente il più spesso trascurato dell'autoreferenza) e deve specificare quali parti vanno stampate normalmente e quali vanno chiuse tra virgolette (o in qualunque altro modo è indicata la materia prima). Da questo punto di vista è carente la frase di Bergerson: anche se distingue chiaramente l'uso dalla menzione affidandosi alle lettere maiuscole per i nomi delle parti elencate e alle lettere minuscole per il conteggio delle parti e le parole di riempitivo, non ha elenchi separati per le parti in maiuscolo e quelle in minuscolo. In questo modo si perde una distinzione essenziale.

Valgono di più, nel Gioco di von Neumann, le soluzioni formulate con parole del vocabolario di base o con seme a livello di lettera (come nella frase di Sallow). La frase di Quine, anche se chiaramente ha in sé un seme (la frase di otto parole racchiusa tra virgolette) e una regola costruttiva (far seguire qualcosa alla propria citazione), non vale perché il suo seme è ben lungi dall'essere materia prima.

C'è un'ottima ragione, tra l'altro, perché il seme della frase di Quine sia così complicato, coincidente in effetti, con la regola costruttiva, a parte le virgolette. La ragione è semplice: bisogna *costruire la regola costruttiva* a partire dalla materia prima e più la regola costruttiva assomiglia al seme più è semplice costruirla partendo dal seme. Per avere una frase totalmente nuova basta fare due copie del seme, effettuare qualche semplice manipolazione per trasformare una copia del seme in regola costruttiva e poi unire l'altra copia del seme alla regola costruttiva appena coniata per ottenere una frase completa nuova di zecca.

Per rendere la cosa più chiara sarà utile fornire una leggera variante alla frase di Quine. Immaginate di saper riconoscere solo le lettere minuscole e che le lettere maiuscole vi siano del tutto estranee. Il testo stampato in lettere maiuscole sarà allora sprovvisto, a tutti gli effetti pratici, di significato e interesse, mentre il testo stampato in minuscolo sarà ricco di un significato e interesse, capace di suggerire idee o azioni. Supponiamo ora che qualcuno vi abbia dato una tavola di conversione che associ a ogni lettera maiuscola la sua corrispondente minuscola, così che possiate «decodificare» il testo in maiuscolo. Un giorno, poi, vi trovate davanti a questo incomprensibile testo in lettere maiuscole:

COMPORTA IL FALSO QUANDO USATO COME SOGGETTO DELLA SUA VERSIONE IN MINUSCOLE

Una volta decodificato, il testo porterebbe a una frase in lettere minuscole, o piuttosto a un frammento di frase in minuscole, un predicato senza soggetto. Suggestivo, eh?

Questa nozione di due alfabeti paralleli, uno in cui il testo è inerte e privo di significato, l'altro in cui il testo è attivo e ricco di significato, non serve solo a costruire una variante di poco conto della frase di Quine; in realtà è molto simile a un trucco di incredibile intelligenza che la natura ha scoperto e sfruttato in ogni cellula di ogni organismo vivente. Il nostro seme, il nostro DNA, è un enorme libro fatto di testi inerti scritti in un alfabeto chimico con 64 lettere «maiuscole» (codoni). Le nostre regole costruttive, i nostri enzimi, sono brevi e concise parole d'ordine scritte in un differente alfabeto chimico con esattamente 20 lettere «minuscole» (amminoacidi). C'è una funzione (il codice genetico) che trasforma le lettere maiuscole in lettere minuscole. Ovviamente alcune lettere minuscole devono corrispondere a più di una lettera maiuscola, ma non è che un dettaglio. Tre caratteri dell'alfabeto maiuscolo, inoltre, non sono lettere, ma segni d'interpunzione che ci dicono dove termina una parola d'ordine e inizia la successiva, ma anche questi sono dettagli. Una volta conosciute le corrispondenze, spesso non ci si ricorderà nemmeno di distinguere tra i due alfabeti chimici: l'alfabeto inerte maiuscolo dei codoni e l'alfabeto attivo minuscolo degli amminoacidi. Ciò che conta è che, armati del codice genetico, si può leggere il libro del DNA (seme) come se fosse una sequenza di parole d'ordine d'enzimi (regole costruttive) che ci dicono come scrivere un nuovo libro di DNA e un nuovo insieme di parole d'ordine d'enzimi. La cosa è perfettamente analoga alla nostra variante della frase di Quine, dove il testo seme, inerte e in maiuscole, era trasformato in testo regola, attivo e in minuscole, che ci diceva come costruire una copia dell'intera frase di Quine una volta dato il seme.

Il DNA di una cellula e gli enzimi si comportano come il seme e le regole costruttive della frase di Quine o come l'elenco delle parti e le regole costruttive dell'automa autoriproducentesi di von Neumann o ancora come il seme e le regole costruttive dei programmi di calcolo che si stampano da sé. È incredibile quanto sia universale il meccanismo dell'autoriferimento e trovo sempre curioso il fatto che la gente che inveisce contro la supposta stupidità dell'autoriferimento sia essa stessa composta da bilioni di molecole autoreferenziali.

Insieme a Scott Kim ho scoperto una strana coppia di frasi:

«La frase seguente è del tutto identica a

questa, tranne che sono state scambiate le parole "seguente" e "precedente", così come le parole "tranne" e "dato" e le espressioni "identica a" e "differente da".»

«La frase precedente è del tutto differente da questa, dato che sono state scambiate le parole "precedente" e "seguente", così come le parole "dato" e "tranne" e le espressioni "differente da" e "identica a".»

A prima vista queste frasi ricordano una variante a due passi del paradosso di Epimenide («La frase seguente è vera»; «La frase precedente è falsa»). Ripensandoci, però, si vede che dicono entrambe esattamente la stessa cosa. Don Byrd non è d'accordo con me: egli sostiene che dicono cose del tutto diverse.

Non sorprende il fatto che molte delle frasi proposte dai lettori avessero qualcosa di paradossale. Alcune erano varianti del paradosso di Russell sul barbiere che rade tutti quelli che non radono se stessi, o sull'insieme di tutti gli insiemi che non includono se stessi come elemento. Per esempio, Gerald Hull ha elaborato questa strana frase: «Questa frase si riferisce a ogni frase che non si riferisce a se stessa». La risposta di Hull è autoreferenziale o no? Sulla stessa linea è la dedica, citata da Michael Gardner, della tesi di uno studente del Reed College: «Questa tesi è dedicata a tutti quelli che non hanno dedicato la loro tesi a se stessi.» Come mi ha fatto notare Charles Brenner, una dedica analoga si trova nel libro *Teoria dei modelli,* di C. C. Chang e H. J. Keisler (trad. it. Boringhieri, 1978). Sempre Brenner suggerisce un'altra variante del paradosso di Russell: elaborare un programma che stampi un elenco di tutti i programmi che non stampano mai se stessi. La domanda, naturalmente, è se questo programma potrà mai stampare se stesso.

Una delle frasi più sconcertanti mi è stata inviata da Robert Boeninger: «Questa frase non ha in realtà la proprietà che sostiene di non avere.» Capito? Il problema, naturalmente, è trovare la proprietà che la frase sostiene di non avere.

Il matematico olandese Hans Freudenthal mi ha inviato un aneddoto paradossale basato sull'autoriferimento:

«C'è una storia di Christian Gellert, un tedesco del XVIII secolo, intitolata "Der Bauer und sein Sohn" ("Il contadino e suo figlio"). Un giorno, durante una passeggiata, il figlio dice una grossa bugia e suo padre lo avverte con severe parole che stanno avvicinandosi al ponte dei bugiardi, un ponte che crolla ogni volta che un bugiardo lo attraversa. Sentito questo terribile avvertimento, il figlio ammette la sua bugia e confessa la verità.

«Quando io [Freudenthal] raccontai questa storia a un bambino di 10 anni, questi mi chiese cosa fosse accaduto quando infine i due erano giunti al ponte. Io risposi: "Crollò sotto il padre, il quale aveva mentito dato che in realtà non esiste un ponte dei bugiardi". (O forse esisteva?)».

C. W. Smith mi ha inviato da London, Ontario, una lettera in cui viene descritta una situazione che ricorda il paradosso di Epimenide:

«Negli anni sessanta, si ergeva solitario in un campo di erbacce in questa città un cartello che diceva: "ricompensa di 25 dollari per informazioni che portino all'arresto e alla condanna di chiunque rimuova questo cartello". Comunque sia, il cartello è scomparso da tempo; e così anche il campo, per altro.»

Detto per inciso, il paradosso di Epimenide non va confuso con il paradosso di Nixonide, coniato per la prima volta da Nixonide il Cretino nel 1974 d.C.: «Questo rapporto non ha valore». A proposito di Epimenide, una delle più belle varianti del suo paradosso è la parte dedicata agli «Errata» in un libro ipotetico descritto da Beverly Rowe:

**(vi)**

**Errata**

**Pagina (vi): Al posto di «Errata» leggere «Erratum»**

Strettamente connesse alle frasi realmente paradossali sono quelle che appartengono alle categorie che io chiamo «nevrotica» e «sana». Una frase sana è una frase che, per così dire, mette in pratica ciò che predica, mentre una frase nevrotica è una frase che dice una cosa e fa l'opposto. Un buon esempio di ciascuna delle due categorie ci è dato da Alan Auerbach. La sua frase sana è «Conciso!», mentre quella nevrotica è «Per scrivere bene - e l'avete sentito milioni di volte - bisogna evitare le esagerazioni.» Una frase sana di Brad Shelton: «Ottantasette parole fa, questa frase non era ancora iniziata». Una delle frasi che più delle altre esce dal sistema è quella di Carl Bender: «La frase seguente è scritta in Tailandia, su».

Consideriamo una frase che ha a che vedere con la precedente e mi è stata inviata da David Stork: «Non c'è bisogno di dire che...». A quale categoria appartiene? Forse è una frase psicotica, come quella inviataci da Pete Maclean: «Se si invertissero i significati di "vero" e "falso", allora questa frase non sarebbe falsa.» Non sono ancora riuscito a capire che cosa significhi! Dan Krimm mi ha scritto: «Ho sentito dire che questa frase è un pettegolezzo.» La seguente è di Linda Simonetti: «Che non è una frase completa, ma solo una proposizione subordinata.» Di Douglas Wolfe è la seguente frase nevrotica basata sull'esperienza: «Non usate mai l'imperativo, e non è neanche corretto costruire una frase mescolando modi diversi.» David Moser mi ha ricordato uno slogan usato un tempo dal *National Lampoon*: «Così divertente che vende senza bisogno di slogan!». Perry Weddle ha scritto: «Sto cercando di insegnare al mio pappagallo a dire "Non capisco nulla di quello che dico". Quando *io* lo dico, è un circolo vizioso autoreferenziale, ma nel *suo* caso?». Stephen Coombs rileva che «una frase può riferirsi a se stessa nel verbo». Mia madre, Nancy Hofstadter, ha sentito il Segretario di stato Haig che descriveva un messaggio d'avvertimento ai Russi come «un'ambiguità calcolata che sarebbe stata chiaramente capita». Sissignore!

Jim Propp propone una sequenza di frasi che vanno dal nevroticamente sano al sanamente nevrotico:

«Questa frase ogni terza, ma essa ancora comprensibile.»

«Questo più facile capire meno omesso.»

«Questo impossibile tranne contesto.»

«4'33" tentativo idea.»

« »

La penultima frase si riferisce al famoso pezzo per pianoforte di John Cage fatto di 4 minuti e 33 secondi di silenzio. L'ultima frase potrebbe essere un brano tratto da *L'intelligenza e la saggezza di Spiro T. Agnew,* anche se è troppo breve per poterne essere certi. Ancora Propp mi ha inviato la seguente citazione sana dal libro *Intelligence in Ape and Man* di David Premack: «Per "produttività" del linguaggio intendo la capacità che il linguaggio ha di introdurre nuove parole in termini di quelle già esistenti.»

Il seguente, inviatomi dal filosofo Howard DeLong, potrebbe essere considerato un sillogismo nevrotico: «Tutti i sillogismi non validi violano almeno una regola. Questo sillogismo viola almeno una regola. Quindi questo sillogismo è un sillogismo non valido.»

Molti lettori hanno rispolverato vecchie frasi e giochi di parole. D. A. Treissman, per esempio, mi fa presente che «la nostalgia non è più come una volta». Henry Taves ricorda quelle deliziose magliette con scritte del tipo «I miei sono andati in Florida e non sono riusciti a portarmi altro che questa schifosa maglietta!» John Fletcher racconta un episodio di una trasmissione televisiva di qualche anno fa in cui Joanne Worley cantava «Sono solo una ragazza che non sa dire "n...", "n...", "n..."» John Healy scrive: «Pensavo di essere indeciso, ma ora non ne sono più così sicuro.»

Anch'io ho da dare qualche contributo a questa raccolta. Un esempio sano è: «In questa frase le ultime tre parole "sono state tralasciate".» O è nevrotico? Queste cose mi confondono! In ogni caso, questa è una frase del tutto sana: «Questa frase offre al(ai) suo(i) lettore(i) varie alternative/opzioni che egli(o essi) è(sono) libero(i) di accettare e/o respingere.» E poi c'è l'inevitabile «Questa frase è nevrotica.» Le cose stanno così: se *è* nevrotica, mette in pratica quel che predica, quindi è sana e *non può* essere nevrotica, ma se è *non* nevrotica, è l'opposto di quello che dice di essere, quindi *deve* essere nevrotica. Non c'è da stupirsi se è nevrotica!

A proposito di frasi nevrotiche, cosa dire delle frasi con crisi di identità? Per me sono le più interessanti di tutte. Un esempio tipico è la domanda, vagamente apprensiva, di Dan Krimm: «Se affermas-

si qualcosa d'altro, si tratterebbe ancora di me?». Non ero ancora soddisfatto e allora ne ho scritta un'altra versione: «In un altro mondo avrei potuto essere una frase su Humphrey Bogart?» Fermatomi a riflettere su quello che avevo fatto, mi accorsi che rivedendo la frase di Krimm avevo alterato la sua identità proprio nel modo che essa temeva. Restava la domanda: tutte queste varianti rimanevano in fondo davvero la stessa frase? Il mio ultimo esperimento in questa direzione: «In un altro mondo, questa frase avrebbe potuto essere la frase di Dan Krimm?»

Alcuni lettori avevano chiaramente seguito questa stessa linea di pensiero. John Atkins chiedeva: «Qualcuno può spiegare perché questa sarebbe ancora la stessa rivista senza questa domanda, mentre questa non sarebbe più la stessa domanda senza questa parola?» Loul McIntosh, che lavora in un centro di riabilitazione per ex schizofrenici, poneva un quesito che connette l'identità personale alle frasi autoreferenziali: «Se io fossi te, chi starebbe leggendo questa frase?» Aggiungeva poi: «(N.B.: Ecco quello che mi succede a lavorare con gli schizofrenici.)» Questo mi porta a Peter M. Brigham, M. D., che nel suo lavoro si era trovato di fronte a un caso di schizofrenia letteraria: «Naturalmente lei ha appena iniziato a leggere la frase che ha appena finito di leggere.»

Un po' di tempo fa ho sentito uno spettacolo radiofonico sugli animali domestici in cui l'annunciatore faceva questa riflessione: «Se un cane avesse scritto questo programma, forse avrebbe detto che gli uomini sono inferiori perché non scondinzolano.» La cosa mi diede da pensare: come sarebbe stato questo articolo se fosse stato scritto da un cane? Non ne sono *sicuro,* ma ho l'impressione che avrebbe avuto a che fare con la caccia agli scoiattoli. Forse avrebbe avuto un paragrafo in cui ci si sarebbe chiesti come sarebbe stato questo articolo se fosse stato scritto da uno scoiattolo.

È di Harold Cooper la frase che più mi è piaciuta tra tutte quelle inviatemi. Gli era stata ispirata dalla mia domanda autoreferenziale controfattuale: «Come sarebbe questa frase se $\pi$ fosse uguale a 3?» Ecco la sua risposta:

> Se $\pi$ fosse uguale a 3, questa frase avrebbe pressapoco quest⬢ aspett⬡

Le O a sei lati rappresentano il fatto che il rapporto tra la circonferenza e il diametro di un esagono è uguale a 3. È chiaro che per Cooper è quanto mai naturale ritenere che se $\pi$ fosse uguale a 3 *i cerchi sarebbero esagoni!* Chi non sarebbe d'accordo? Mi colpì il fatto che, portando a 3 il valore di $\pi$, non solo i cerchi si trasformavano in esagoni, ma anche le frasi interrogative si tramutavano in frasi dichiarative. Si ricordi che la domanda chiedeva come la domanda stessa sarebbe stata in quello strano mondo congiuntivo. Avrebbe perso la curiosità su se stessa e avrebbe cessato di essere una domanda? Non vedevo perché quel tratto della personalità della frase avrebbe dovuto essere influenzato dal valore di $\pi$. Del resto mi sembrava ovvio che, se $\pi$ *fosse* uguale a 3, l'antecedente del condizionale non sarebbe più congiuntivo. In effetti, invece che «se $\pi$ *fosse* uguale a 3», dovrebbe essere «dato che $\pi$ *è* uguale a 3» (o qualcosa del genere). Ricapitolando quello che avevo pensato, costruii una variante della frase di Cooper: «Com'è questa frase, essendo $\pi$ uguale a 3 (come sempre)?»

Molti lettori sono apparsi interessati alle frasi che fanno riferimento al linguaggio che si sta usando (o non si sta usando, come può accadere). Un esempio è: «Se parlaste italiano, parlereste la vostra lingua nazionale.» Jim Propp mi ha inviato una deliziosa coppia di frasi di questo tipo, che vanno lette insieme:

«Cette phrase se refere a elle-meme, mais d'une maniere peu evidente a la plupart des Italiens».

«Plim glorkle pegram ut replat, trull gen ris clanter froat veb nup lamerack gla smurp Terrestri.»

Se non capite la prima frase, fatevi aiutare da un amico marziano a decodificare la seconda: vi darà dei suggerimenti sulla prima. (Ci scusiamo per aver saltato i corretti accenti marziani, ma non avevamo i caratteri tipografici adatti.)

Lo scorso aprile pubblicai numerose frasi create da David Moser e affermai che egli aveva scritto un'intera storia formata da frasi autoreferenziali. Molti lettori si sono dimostrati interessati alla cosa e ho deciso che non c'era miglior modo per concludere quest'articolo che riprodurre interamente la storia di David. Eccola:

Questo è il titolo di questa storia,
che si trova anche parecchie volte
nella storia stessa

Questa è la prima frase di questa storia. Questa è la seconda frase. Questo è il titolo di questa storia, che si trova anche parecchie volte nella storia stessa. Questa frase si interroga sul valore intrinseco delle prime due frasi. Questa frase vi informa, nel caso non ve ne foste già accorti, che questa è una storia autoreferenziale, cioè una storia formata da frasi che si riferiscono alla loro propria struttura e funzione. Questa è una frase che serve per finire il primo capoverso.

Questa è la prima frase di un nuovo capoverso di una storia autoreferenziale. Questa frase vi fa conoscere il protagonista della storia, un ragazzo di nome Billy. Questa frase vi dice che Billy è biondo, con gli occhi azzurri, americano, ha 12 anni e strangola sua madre. Questa frase fa un commento sulla goffa natura della forma narrativa autoreferenziale mentre riconosce lo strano e divertente distacco che garantisce allo scrittore. Come per illustrare quanto detto nell'ultima frase, questa frase ci ricorda, senza ombra di facezia, che i bambini sono un prezioso dono divino e che il mondo è un posto migliore quando è benedetto dalle gioie e delizie che essi vi portano.

Questa frase descrive gli occhi della madre di Billy che schizzano fuori dalle orbite e la sua lingua che penzola e fa riferimento agli spiacevoli gorgoglii che sta emettendo. Questa frase fa l'osservazione che questi sono tempi incerti e difficili e che i rapporti, anche quelli apparentemente permanenti e profondamente radicati, hanno la tendenza a rompersi.

Introduce in questo capoverso l'espediente dei frammenti di frase. Un frammento di frase. Un altro. Buon espediente. Sarà più usato in seguito.

Questa è in realtà l'ultima frase della storia, ma è stata messa qui per sbaglio. Questo è il titolo di questa storia, che si trova anche parecchie volte nella storia stessa. Svegliatosi una mattina uscendo da sgradevoli sogni, Gregor Samsa si trovò nel suo letto trasformato in un gigantesco insetto. Questa frase vi informa che la frase precedente è di un'altra storia (una molto migliore, va detto) e non c'entra niente in questo passo della narrazione. A dispetto delle affermazioni della frase precedente, questa frase si sente obbligata a informarvi che la storia che state leggendo è in realtà *La metamorfosi,* di Franz Kafka, e che la frase a cui ci si riferiva nella frase precedente è l'unica frase che appartiene davvero a questa storia. Questa frase annulla la frase precedente informando il lettore (povero, confuso infelice) che questo brano letterario è in realtà la Dichiarazione di indipendenza, ma che l'autore, in un gesto di estrema negligenza (se non di malizioso sabotaggio), non ha finora incluso nemmeno *una sola frase* di quel commovente documento, anche se ha accondisceso a usare un piccolo *frammento* di frase, vale a dire «quando nel corso degli umani eventi», chiuso tra virgolette vicino alla fine della frase. Mostrando una chiara consapevolezza della noia e della sincera ostilità del lettore medio verso gli inutili giochi concettuali a cui si indulgeva nelle frasi precedenti, *questa* frase ci riporta infine allo scenario della storia ponendo la domanda: «Perché Billy strangola sua madre?» Questa frase cerca di fare un po' di luce sulla domanda posta dalla frase precedente, ma non ci riesce. *Questa* frase, invece, ci riesce suggerendo una possibile relazione incestuosa tra Billy e sua madre e alludendo alle concomitanti complicazioni freudiane che ogni lettore astuto vedrà immediatamente. Incesto. L'innominabile tabù. La proibizione universale. Incesto. E si notano i frammenti di frase? Buon espediente letterario. Sarà più usato in seguito.

Questa è la prima frase di un nuovo capoverso. Questa è l'ultima frase di un nuovo capoverso.

Questa frase può servire sia da inizio di un capoverso sia da fine, a seconda della sua posizione. Questo è il titolo di questa storia, che si trova anche parecchie volte nella storia stessa. Questa frase solleva una seria obiezione verso l'intera classe di frasi autoreferenziali che si limitano a commentare la propria funzione o la pro-

pria posizione all'interno della storia (ad esempio le ultime quattro frasi), per il fatto che esse sono monotonamente prevedibili, indimenticabilmente autoindulgenti e servono semplicemente a distrarre il lettore dal vero argomento di questa storia, che a questo punto sembra avere a che fare con strangolamento e incesto e chissà quali altre delizie. Lo scopo di questa frase è rilevare che la frase precedente, anche se non appartiene essa stessa alla classe di frasi autoreferenziali contro cui solleva obiezioni, nondimeno serve *anch'essa* semplicemente a distrarre il lettore dal vero argomento di questa storia, che in realtà ha a che fare con l'inesplicabile trasformazione di Gregor Samsa in un insetto gigantesco (a dispetto delle rumorose affermazioni contrarie di altre benintenzionate, ma male informate frasi). Questa frase può servire sia da inizio di un capoverso sia da fine, a seconda della sua posizione.

Questo è il titolo di questa storia, che si trova anche parecchie volte nella storia stessa. Questo è *quasi* il titolo di questa storia, che si trova solo una volta nella storia stessa. Questa frase afferma a malincuore che fino a questo punto il modo narrativo autoreferenziale ha avuto un effetto paralizzante sul progresso della storia stessa, cioè queste frasi sono state talmente impegnate ad analizzare se stesse e il loro ruolo nella storia che hanno finito per non compiere la loro funzione di comunicatori di eventi e idee che uno spera si fondano in una trama, in uno sviluppo dei personaggi, ecc., in breve, la vera *raison d'être* di ogni rispettabile frase all'interno di un brano di irresistibile prosa. Questa frase rileva inoltre l'ovvia analogia tra la condizione di queste agonizzanti frasi autocoscienti e le analoghe afflizioni degli esseri umani, e sottolinea gli analoghi effetti paralizzanti che scaturiscono da un'eccessiva e torturata autoanalisi.

Lo scopo di questa frase (che può anche servire da capoverso) è di riflettere che se la Dichiarazione di indipendenza fosse stata formulata e strutturata allo stesso modo lezioso e incoerente di questa storia, non si potrebbe dire in che tipo di corrotta società libertina vivremmo ora o in quali abissi di decadenza potrebbero essere precipitati gli abitanti di questo paese, fino al punto che scrittori squilibrati e degradati costruiscano frasi irritantemente ingombranti e inutilmente prolisse che a volte possiedono la discutibile se non del tutto indesiderabile proprietà di riferirsi a se stesse ed esse a volte arrivano a divenire frasi sconclusionate o mostrano altri segni di inescusabile sciatteria grammaticale quali non necessarie ridondanze superflue che quasi certamente avrebbero insidiosi effetti sullo stile di vita e sulla morale della nostra impressionabile gioventù, portandola a commettere incesto o addirittura assassinio e forse *è per questo* che Billy sta strangolando sua madre, a causa di frasi *proprio come questa*, che non hanno obiettivi discernibili o scopi perspicui e terminano in un punto qualsiasi, anche in mez

Bizzarro. Un frammento di frase. Un altro frammento. Dodici anni. Questa è una frase che. Frammentata. E strangolando sua madre. Scusate, scusate. Bizzarro. Questo. Più frammenti. Questo è. Frammenti. Il titolo di questa storia, che. Biondo. Scusate, scusate. Frammento dopo frammento. Più difficile. Questa è una frase che. Frammenti. Diavolo buon espediente.

Lo scopo di questa frase è triplice: (1) scusarsi per la sfortunata e inesplicabile scorrettezza nel capoverso precedente; (2) assicurarti, lettore, che non accadrà più; e (3) ripetere l'affermazione che questi sono tempi incerti e difficili e che aspetti del linguaggio, anche quelli apparentemente stabili e profondamente radicati come la sintassi e il significato, si spezzano. Questa frase non aggiunge nulla di sostanziale a ciò che è stato espresso nella frase precedente, ma serve semplicemente da frase conclusiva per questo capoverso, che altrimenti potrebbe non averne una.

Questa frase, in un improvviso e coraggioso soprassalto di altruismo, cerca di abbandonare il modo autoreferenziale, ma fallisce. Questa frase cerca di nuovo, ma il tentativo è condannato in partenza.

Questa frase, in un ultimo e disperato tentativo di infondere un minimo di narrazione in questa prosa paralizzata, allude rapidamente ai frenetici tentativi di Billy di nascondersi, seguita da un lirico, toccante e bel passaggio in cui Billy si riconcilia con suo padre (risolvendo così i subliminali conflitti freudiani che risultano ovvi a qualsiasi astuto lettore) e da un'eccitante scena finale di caccia da parte della polizia durante la quale Billy è accidentalmente colpito e ucciso da una recluta della polizia in preda al panico che per coincidenza si chiama anch'essa Billy. Questa frase, anche se fondamentalmente in completa simpatia con i lodevoli sforzi dell'ultima frase densa d'azione, ricorda al lettore che queste allusioni a una storia che in realtà non esiste ancora non sono in alcun modo sostituti della cosa reale e pertanto non possono tirar fuori l'autore (quello stupido indolente) dalla proverbiale trappola.

Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso. Capoverso.

Lo scopo. Di questo capoverso. È di scusarsi. Per l'uso gratuito. Di. Frammenti di frase. Scusate.

Lo scopo di questa frase è scusarsi per gli inutili e sciocchi giochetti da adolescenti a cui si indulgeva nei due precedenti capoverso, ed esprimere da parte di noi, le frasi più mature, il rincrescimento per il fatto che, nel complesso, il tono di questa storia è tale da non riuscire a comunicare un semplice, per quanto un po' torbido, scenario.

Questa frase desidera scusarsi per le inutili scuse che si trovano in questa storia (questa inclusa), che, anche se manifestamente qui poste a beneficio dei lettori più irritati, non fanno che ritardare in modo esasperantemente ripetitivo la prosecuzione dell'ormai quasi dimenticata linea narrativa.

Questa frase mette l'accento sul terribile significato dell'autoriferimento applicato alle frasi, una pratica che potrebbe dimostrarsi un vero vaso di Pandora di potenziale rovina, in quanto se una frase può riferirsi o alludere a se stessa, perché non un'umile proposizione subordinata, magari *proprio questa?* O questo frammento di frase? O tre parole? Due parole? *Una?*

Forse è giusto che questa frase ci ricordi gentilmente e senza ombra di condiscendenza che questi sono davvero tempi difficili e incerti e che in genere le persone non sono buone l'una con l'altra, e che forse noi, esseri umani senzienti o frasi senzienti che siamo, dovremmo *sforzarci di più*. Voglio dire che *c'è* qualcosa come il libero arbitrio, ci *deve* essere, e questa frase ne è la prova! Né questa frase né tu, lettore, siete completamente senza speranza davanti alle impietose forze che operano nell'universo. Dovremmo stare ben piantati per terra, affrontare i fatti, prendere Madre Natura per il collo e *sforzarci di più*. Per la gola. Di più. Di più, di più.

Scusate.

Questa è l'ultima frase della storia. Questa è l'ultima frase della storia. Questa è l'ultima frase della storia. Questa è.

Scusate.